

## Article

# Development of a Multilingual Model for Machine Sentiment Analysis in the Serbian Language

Drazen Draskovic \*, Darinka Zecevic and Bosko Nikolic 

Department of Computer Science and Information Technology, School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia

\* Correspondence: drazen.draskovic@etf.bg.ac.rs

**Abstract:** In this research, a method of developing a machine model for sentiment processing in the Serbian language is presented. The Serbian language, unlike English and other popular languages, belongs to the group of languages with limited resources. Three different data sets were used as a data source: a balanced set of music album reviews, a balanced set of movie reviews, and a balanced set of music album reviews in English—MARD—which was translated into Serbian. The evaluation included applying developed models with three standard algorithms for classification problems (naive Bayes, logistic regression, and support vector machine) and applying a hybrid model, which produced the best results. The models were trained on each of the three data sets, while a set of music reviews originally written in Serbian was used for testing the model. By comparing the results of the developed model, the possibility of expanding the data set for the development of the machine model was also evaluated.

**Keywords:** sentiment analysis; low-resource languages; ML-based tools for NLP; Serbian language



**Citation:** Draskovic, D.; Zecevic, D.; Nikolic, B. Development of a Multilingual Model for Machine Sentiment Analysis in the Serbian Language. *Mathematics* **2022**, *10*, 3236. <https://doi.org/10.3390/math10183236>

Academic Editors: Nebojsa Bacanin, Catalin Stoean and Florin Leon

Received: 18 July 2022

Accepted: 3 September 2022

Published: 6 September 2022

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



**Copyright:** © 2022 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

One of the important subfields of machine learning is natural language processing (NLP). It includes the development of software systems that are able to automatically analyze and understand natural human languages. The largest amount of research in the field of natural language processing has been done for the English language, which is the most widely spoken language in the world.

Sentiment analysis, as a segment in natural language processing, deals with the development of models that are able to determine the subject's attitude on a given topic based on the content of the text [1]. The most common example is surveying users' opinions on forums, news portals, online stores, social networks, etc. [2–6].

This research aims to develop a machine model for sentiment analysis in the Serbian language. The creation of a model for the Serbian language is largely conditioned by the availability of resources, and for this reason, the first part of the research dealt exclusively with the collection of data for machine processing. During the development of the model, some available data sets in the Serbian language were used, as well as data sets in the English language, which were translated into the Serbian language using the Google Translate API.

In today's NLP research, English is still the most dominant language [7]. The cause of such a situation is the important use of the English language in international communication, and it is a consequence of the development of many new NLP solutions for the English-speaking area. In addition, the public availability of NLP resources for the English language is very high, which increases the amount of research on that language [8]. In the digital sphere, it is estimated that 60% of the content on the Internet is written in English.

With the development of NLP models for deep learning, the number of data sets of the most popular languages is increasing because such models have to work with large amounts of training data. On the other hand, collecting a large amount of data and data sets for fewer languages is very difficult. Recently, the interest of researchers in the computer processing of other languages, such as Chinese, Japanese, German, or Arabic, has been increasing [9,10]. This trend can be seen in [11], where the most-represented natural languages in NLP research over the last 20 years are shown, and at the top are English, Mandarin Chinese, Japanese, and German, followed by Arabic, French, Spanish, Italian, and Czech. Currently, it is considered that there are approximately 7100 languages on the planet, of which 4000 have a script. Still, more than 6500 languages are not present in digital form, i.e., it is impossible to find data sets in those languages on the Internet [12].

A precise definition of low-resource languages does not exist in NLP research [7,13–15]. In this study, the Serbian language will be considered as a language with low resources because it meets the following criteria: (a) the Serbian language is moderately widespread in digital form; (b) there are a small number of available preprocessing tools and resources in Serbian, but it is not possible to obtain large amounts of annotated data in Serbian for the creation of NLP models; and (c) the number of researchers involved in the development of NLP tools for the Serbian language is minimal and financial resources are limited. Furthermore, applications of a multilingual or cross-lingual approach are rare in the Serbian language, as is the development of resources for analysis. This was also the main motivation of the authors, as they intend to use the realized models to show what would be best in the sentiment analysis of short texts.

The second section provides an overview of the application of multilingual models based on traditional machine learning algorithms and modern approaches. The third section describes the methods used for collecting and preparing the data for machine processing. It explains some of the most important algorithms used for sentiment analysis, as well as the methodologies applied for extracting the useful attributes from the text. In the fourth section, the results of developing of the model for the Serbian language are presented. Discussion is presented in the fifth section. Finally, the last section describes the key outcomes of the research and provides suggestions for improving the sentiment analysis model for the Serbian language.

## 2. Related Work

Sentiment analysis (or opinion mining) represents the problem of the automatic detection and processing of attitudes, assessments, and opinions expressed by people towards certain entities, persons, events, issues, topics, and their properties [16,17]. The main data sources for sentiment analysis are blogs and entertainment sites with user reviews, e-commerce sites such as Amazon, eBay, etc. with user reviews, social media content generated on Twitter and Facebook, and data from communication mediums such as SMS, WhatsApp, Viber, etc.

The basic categories of sentiment analysis techniques are lexicon and machine learning techniques. In the first group, opinions are identified based on manual or automatic processing techniques, such as dictionary-based or corpus-based methods. Most machine learning techniques consider sentiment analysis as a supervised learning problem, but modern methods have also explored semi-supervised approaches. Sentiment analysis is usually classified as a classification problem [18]. This problem includes narrower problems such as polarity detection, where the goal is the binary division of texts into positive and negative. Further, this problem contains subjectivity text detection, where the goal is to distinguish objective texts from subjective ones. Sarcasm detection is also represented in the research as a sentiment analysis problem [19]. Due to the enormous commercial need for the automatic detection and processing of people's attitudes, sentiment analysis represents an NLP problem with one of the most practical applications.

A multilingual model is a single model that can handle multiple languages simultaneously. The biggest problem with sentiment analysis research is the lack of resources

for low-resource languages, such as Serbian [20]. In a study on the needs of a developed machine model, reviews in Serbian and English were used. For the efficient sentiment analysis of multilingual content and the identification of positive and negative comments, the authors used traditional techniques such as a naive Bayes classifier, SVM classifier, simple neural networks, convolutional neural networks, and recurrent neural networks—the most famous of which is long short-term memory (LSTM) [21,22]. Most of today's researchers use powerful models such as mBERT (multilingual bidirectional encoder representations from transformers), which supports 104 languages [23–25]. In some research, an LSTM model was used instead of the BERT model to capture the sentiment of multilingual comments [26,27]. For example, the research by Žitnik et al. [28] applied a customized BERT adapter to a newly annotated data set of Slovene news articles. Other authors consider an Electra approach to be computationally more efficient than a BERT model, and the authors of [29] developed a transformer model that was pre-trained on 8 billion tokens of crawled text from web domains with South Slavic languages.

Mozetič et al. [30] conducted multilingual sentiment analysis for 13 languages, including Serbian, Bosnian, Croatian, Slovenian, Bulgarian, Slovak, etc., based on Twitter data and compared the performance of the most famous classifier models. They concluded that the size and quality of the data sets impacts the performance more than the model selection does. The rules of negation in the Serbian language and their influence on polarity in Twitter data were investigated by Ljajić and Marovac [31]. They used a lexicon-based approach and machine learning methods. Batanović dealt with the sentiment analysis and semantic similarity of short texts in the Serbian language [32].

Some authors have used IMDb user movie reviews and translated them into Serbian, receiving outstanding results [33]. Other authors have translated tweets written in various European languages into English to overlook the results of the Eurovision song contest [34]. The authors of [35] presented the process of developing a sentiment analysis framework for the Serbian language. Stankovic et al. presented a study on the sentiment analysis of Serbian novels from the period 1840–1920. Their comparison shows that models trained on the labeled data sets of movie reviews indicate that they cannot successfully be used for the sentiment analysis of sentences in old novels [36].

Table 1 shows a review of analyzed research papers in the form of covered languages, applied techniques, and analyzed data sets. The analysis includes data sets in the Serbian language or another low-resource language and different types of sentiment analysis applications in those languages. It can be noticed that research works and open data sets for sentiment analysis in the Serbian language are very rare.

**Table 1.** Related research papers overview.

Author, Year, Language	Data Set	Major Contribution	Techniques
Mozetič et al., 2016, 13 languages, including Slovenian, Serbian, Albanian, Bulgarian, etc. [30]	Twitter data	evaluation of data sets using different classifiers and comparative analysis for multiple languages	NB, different types of SVM
Mladenović et al., 2016, Serbian [35]	movie reviews, news set	building a sentiment analysis framework for Serbian	Maximum Entropy
Ljajić and Marovac, 2018, Serbian [31]	Twitter data	examining how the treatment of negation impacts the sentiment of tweets	NB, LR, SVM, J48-DTree
Lohar et al., 2019, English => Serbian [33]	large movie review data set (Maas, 2011)	building a machine translation system for user-generated content	Moses MT toolkit, OpenNMT
Batanović, 2021, Serbian [32]	movie reviews, book reviews	evaluation and determination of the optimal configurations using several different kinds of machine-learning models on a range of sentiment classification tasks	MNB, CNB, LR, SVM, NB-SVM
Stanković et al., 2022, Serbian [36]	SrpELTeC1 (multilingual corpus of novels)	development and application of sentiment lexicon, (sentence) data set labeling, and training of the models for sentiment analysis	LR, NB, DTree, RF, SVN, k-NN

### 3. Materials and Methods

The most common problem when analyzing sentiment in Serbian and other low-resource languages is the irregular distribution of positive, neutral, and negative examples within one categorized data set. The first example of a balanced data set in the Serbian language is SerbMR, which is based on movie reviews. Based on this data set, a data set was created for sentiment analysis in the Serbian language, containing reviews of music albums and songs.

#### 3.1. Data Sources and Challenges

Since there is no single database with a sufficiently large number of reviews of music albums in the Serbian language, various internet portals were used as data sources. A greater number of sources brings greater diversity in review writing style and review content, evaluation method, etc. All of this complicates the process of formatting the output data in a unique way, and in some cases, it requires manual data processing.

The collection of reviews was done with the help of a developed intelligent agent in the Python programming language using the BeautifulSoup library. For each portal, links with review texts were first collected, and then the text of the review with the grade was separately extracted from the web portal. During the data collection process, there were several key challenges that were classified into several categories:

- absence of negative reviews
- absence of grades with the text of the review itself
- unfavorable web structure of the portal (album reviews were not separated into different categories on the portal and web pages with texts containing reviews could not be automatically filtered from other articles)
- adverse web structure of the web portal with review, including:
  - the grade is not separated from the rest of the text (often in the middle of the text)
  - textual content within the element, with a rating
  - template content at the beginning and/or end of the text
  - unnecessary content with the review itself (for example, JavaScript code)
- different scales and assessment methods

Sources that did not have ratings with the text of the review or that did not have negative reviews in the set were excluded from consideration. Additional programmatic and manual text filtering solved the problems arising from a site's unfavorable structure and pages. Since most portals used a ten-class rating scale where 1 was the lowest rating and 10 was the highest, the other scales were mapped to the ten-class scale. The final data set with the distribution of grades is shown in Table 2.

**Table 2.** Summary overview of the collected reviews.

Web Portal	1	2	3	4	5	6	7	8	9	10	Sum
<a href="#">2kokice.com</a> (accessed on 12 September 2021)	4	0	2	3	0	10	8	26	12	11	76
<a href="#">balkanrock.com</a> (accessed on 13 September 2021)	8	8	17	36	16	95	137	109	60	33	519
<a href="#">popboks.com</a> (accessed on 10 September 2021)	8	13	38	98	205	350	467	295	66	14	1554
<a href="#">serbian-metal.org</a> (accessed on 16 September 2021)	0	0	0	4	4	18	52	108	53	1	240
<a href="#">hardwiredmagazine.com</a> (accessed on 18 September 2021)	0	1	8	2	41	20	1	87	12	17	189
<a href="#">nocturno.com</a> (accessed on 13 September 2021)	3	0	0	3	1	27	42	120	45	40	281
<a href="#">hellycherry.com</a> (accessed on 15 September 2021)	4	2	2	0	1	2	8	5	2	7	33

**Table 2.** Cont.

Web Portal	1	2	3	4	5	6	7	8	9	10	Sum
<a href="#">mnsblog.weebly.com</a> (accessed on 19 September 2021)	3	0	1	2	0	2	4	3	2	3	20
<a href="#">tegla.rs</a> (accessed on 13 September 2021)	165	35	18	44	60	204	270	63	5	34	898
<a href="#">plejer.net</a> (accessed on 15 September 2021)	0	2	0	5	2	16	7	17	10	11	70
<a href="#">Balkanmetalpromotion</a> (accessed on 16 September 2021)	0	0	1	1	0	2	2	7	4	0	17
<a href="#">petar-kostic.blogspot</a> (accessed on 12 September 2021)	0	2	0	10	0	11	0	32	0	2	57
<a href="#">Mislitemojomglavom</a> (accessed on 15 September 2021)	9	5	11	20	13	30	41	123	82	29	363

Table 3 shows the statistics of the collected reviews from the 13 selected web portals. A balanced data set is a set in which the examples of each class are evenly represented. In this data set of music reviews, three classes were formed: negative reviews (grades from 1 to 4), neutral reviews (grades 5 and 6), and positive reviews (grades from 7 to 10). As the number of negative reviews among the collected data was less than the positive and neutral reviews, for each negative review, the best positive and negative pair was searched according to the modification of the algorithm shown in [37]. When pairs were found, the following characteristics of each review were considered:

1. Review grading—negative reviews were paired with positive ones according to the principle of inverse grades, e.g., 1 by 10, 2 by 9, etc., and the subset of neutral reviews consisted of an equal number of reviews rated as 5 and 6.
2. Review length—the difference in the word counts between the pairs should be minimal.
3. Review source—different portals had different review writing styles and different criteria, and they covered different music genres. When pairs were found, preference was given to the reviews from the same portal.

**Table 3.** Statistical presentation of the collected reviews.

Web Portal	Genre	Grade Scale	Number of Reviews	Number of Positives	Number of Neutrals	Number of Negatives	Average Review Length (Words)	Shortest Review (Words)	Longest Review (Words)
2kokice.com	pop	1–10	76	75%	13.2%	12.8%	248	38	612
balkanrock.com	rock, metal, punk	1–10	519	65.3%	21.4%	13.3%	256	47	2072
popboks.com	rock, pop	1–10	1554	54.3%	35.7%	10%	555	73	1915
serbian-metal.org	metal, rock	1–100	240	89%	9%	2%	448	68	1192
hardwiredmagazine.com	rock	1–5	189	62%	32%	6%	517	59	1263
nocturno.com	rock	1–10	281	88%	10%	2%	566	69	1242
hellycherry.com	rock	1–5	33	67%	9%	24%	427	45	1031
mnsblog.weebly.com	pop	1–10	20	60%	10%	30%	464	36	1646
tegla.rs	different	0–5	898	70%	8%	22%	60	1	309
plejer.net	rock	1–5	70	64%	26%	10%	500	58	892
balkanmetalpromotion	rock, metal	1–100	17	76%	12%	12%	590	367	1038
petar-kostic.blogspot	rock	1–5	57	60%	19%	21%	779	399	1472
mislitemojomglavom	different	1–10	363	77%	11%	12%	601	44	1638

The algorithm used consists of the following steps:

1. Finding all potential pairs—for each negative review, a list of possible pairs is found from the positive and neutral set, respecting the above criteria. In case there is no such pair, the first criterion to be relaxed is the source of the review, followed by the differences in the length of the reviews. The criterion of the review score is never relaxed. The criteria are relaxed cyclically until a compatible review is found.
2. Sort negative reviews in ascending order in potential pairs to maximize the number of pairs found in one iteration.

3. Matching of reviews—in case there is a large number of positive candidates, the one with the smallest difference in length is chosen, as is the one that reduces the total difference in length between the positive and negative reviews. Neutral reviews are selected in a similar way, except that the rule of equal representation of reviews with a rating of 5 and 6 is respected as much as possible.

The steps are repeated cyclically until a positive and neutral pair is found for each negative review. An overview of the characteristics of the data set obtained by the used algorithm is given in Table 4 (column A).

**Table 4.** Statistical representation of the music reviews in a balanced set (A), movie reviews (B), and translated reviews (C).

	(A)	(B)	(C)
Total number of reviews	1830	2523	51 234
Number of reviews per class	610	841	17 078
Longest positive review	2025 words	1813 words	2129 words
Longest neutral review	1552 words	1621 words	3125 words
Longest negative review	1664 words	1835 words	1845 words
Shortest positive review	8 words	21 words	1 word
Shortest neutral review	6 words	73 words	2 words
Shortest negative review	1 word	21 words	1 word
Average positive review	489 words	472 words	112 words
Average neutral review	344 words	468 words	132 words
Average negative review	344 words	467 words	101 words

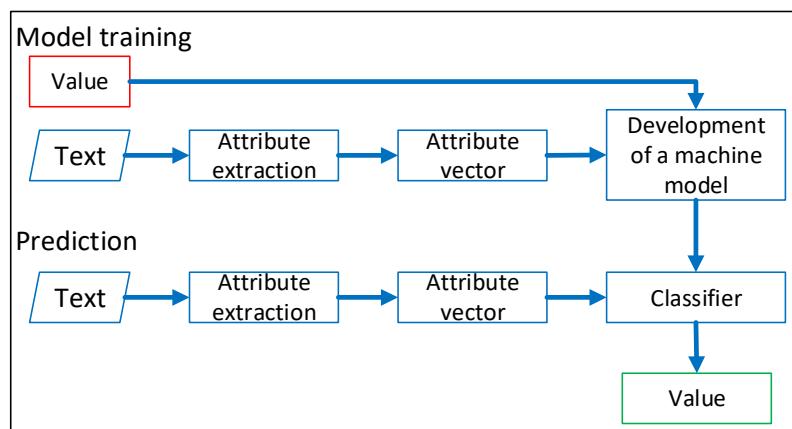
In addition to the described data set of reviews, two more data sets were used in the research because the Serbian language is a languages with limited resources. For the development of the model to be successful, it was necessary that the data be related, and so movie reviews written in Serbian and music album reviews—MARD, originally written in English—were used [38]. The second data set was translated into Serbian using the Google Translate API, and it was used in the model.

In this way, the possibility of expanding the data set for machine text processing was examined. One set was the use of data in the same language, but from a different domain, while the other was the use of data originally written in other languages and translated into Serbian. An overview of the characteristics of the additional two data sets can be found in Table 4 (columns B and C).

### 3.2. Sentiment Analysis

Text sentiment analysis is a subgroup within the text classification process. Based on the content of the text, it is necessary to determine the feelings and the opinions of the author of the text according to the topic described in the text. Examples are hotel reviews, movie reviews, comments on social media, and comments in newspaper articles. Sentiment analysis, as a part of natural language processing, solves two problems: the classification of subjectivity and the classification of polarity. It is necessary to separate the subjective from the objective, as well as the positive from the negative, when expressing an attitude about an entity.

Different natural language processing methods and algorithms are used to determine the sentiment of text. These methods can be divided into manual, automatic, and hybrid. One of the main difficulties in language analysis is the complexity of linguistic expressions, along with morphological forms, irony, metaphors, negation, ambiguity of sentences, etc. A diagram of the machine learning classifier is given in Figure 1.



**Figure 1.** Diagram of a machine classifier.

In manual methods, the sentiment is determined based on some simple rules—the sentiment dictionary. However, manual methods are very naive and unreliable because they do not consider how words are connected in sentences. With automatic methods, determining the sentiment of a text is presented as a classification problem—the input data is observed as a vector of values, and during model training, a function is found that maps that vector to the appropriate class. In the testing process, i.e., the prediction, an attribute vector is created from the input text, which is then passed to the machine model to determine the class based on the selected mapping function.

### 3.3. Text Attribute Vector

The first step in the text sentiment analysis process is to transform the text into an attribute vector. The most frequently used attributes are the presence and frequency of words, the sentiment of words and phrases, and negation.

In the bag-of-words model, text is observed as the unordered set of words contained within it. Each word represents one attribute in the classification model, the order and relationship of the words are ignored, and the value of the attribute is either the number of repetitions of the given word in the text or a binary value (0 or 1) representing the absence or presence of that given word in the text. In addition, sequences of several consecutive words—a bag of n-grams—can be considered as an attribute.

Weighting implies a methodology for determining the importance of a word in a document or set of documents. Types of weighting include term frequency weighting, inverse document frequency weighting, and term frequency-inverse document frequency weighting.

With term frequency (*TF*) weighting, the relevance of document *d* for a specific query increases with the higher frequency of occurrence of the word *t* from the query in the document—not linearly, but logarithmically—as follows:

$$TF = \begin{cases} 1 + \log_{10} Count(t)_d, & Count(t)_d > 0 \\ 0, & Count(t)_d = 0 \end{cases} \quad (1)$$

For the inverse document frequency (*IDF*) technique, words that appear in all documents are less important than words that appear in a small number of documents. In this case, words that rarely occur are given more weight, as follows:

$$IDF = \log_{10} \frac{N}{df_t}, \quad (2)$$

where *N* is the total number of documents in the set and *df<sub>t</sub>* is the number of documents in which the word *t* occurs.

*TF-IDF* (term frequency-inverse document frequency) is a statistical measure that evaluates how relevant a word is to a document in a collection of documents, which is determined as follows:

$$TF - IDF = (1 + \log_{10} Count(t)_d) \cdot \log_{10} \frac{N}{df_t} \quad (3)$$

The Serbian language is a morphologically rich language, and so one word can have many different forms. A word can change by case, gender, singular or plural form, and verb tense. As computer systems cannot recognize fundamentally different words from morphological variations, it is necessary to reduce the various forms to the basic word form. Morphological changes can be classified into two groups:

- Inflectional morphology—different forms of one word (e.g., book, books, etc.)
- Derivational morphology—derivation of new words from the basic one:
  1. derivation by adding a suffix (e.g., logic => logical)
  2. derivation by adding a prefix (e.g., pure => impure)
  3. derivation by combining several words (compounds) (e.g., snowball, grandmother, upstream, etc.)

The two basic methods of morphological normalization are: word stemming and word lemmatization. Stemming is a methodology similar to word rooting, but without the knowledge of linguistics. Stemmers cut off the ends of words but do not recognize the concept of suffixes, and the cutting is instead implemented based on a list of rules (maps or regular expressions).

Lemmatization is a more complex procedure than stemming. It is most often implemented in the form of a separate machine model with the help of morphological dictionaries, which map different forms of words into lemmas. Lemmatization also depends on the context of the text.

### 3.4. Supervised Classification Algorithms

In this research, three standard algorithms were used for the classification problems: naive Bayes (NB), logistic regression (LR), and support vector machine (SVM).

#### 3.4.1. Naive Bayes

A naive Bayes is one of the most widespread and successful algorithms for text classification. The algorithm is based on Bayes' probability theorem:

$$P(y|x) = \frac{P(y) * P(x|y)}{P(x)} \quad (4)$$

where  $y$  is the class,  $x$  is the input data,  $P(y|x)$  is a posterior probability, i.e., the probability that  $y$  will happen if  $x$  happens,  $P(x|y)$  is the certainty function, and  $P(x)$  and  $P(y)$  are the probabilities that  $x$  and  $y$  will occur.

The classification decision is made based on the maximum value of the posterior probability. Based on the input data, we calculate the probability for each of the possible classes and choose the maximum value.

The prefix “naive” comes from the assumption that the attributes are conditionally independent of each other and that all attributes are equally important. Although the assumptions about attribute independence are often not correct, in practice, this classifier offers good results.

### 3.4.2. Logistic Regression

Logistic regression is the use of a linear regression model for a classification problem. The logistic regression model belongs to the probabilistic classifier. The logistic regression hypothesis is:

$$h(x) = \frac{1}{1 + e^{-(\omega_0 + \omega_1 x_1 + \dots + \omega_n x_n)}} \quad (5)$$

where  $n$  is the number of features used in the model. The new data is added to the class that is more probable for it. For  $h(x) > 0.5$ , the data is classified in the class  $y = 1$ , and for  $h(x) < 0.5$ , it is in the class  $y = 0$ .

By introducing the fictitious feature  $x_0 = 1$ , the hypothesis is transformed into the following form:

$$h(x) = \frac{e^{W^T X}}{e^{W^T X} + 1} \quad (6)$$

where  $W$  represents the vector of all weight parameters,  $X$  represents the vector of all attribute values, and  $W^T X$  is their scalar product. Then, for the class separation hyperplane  $h(x) = 0.5$ , the following applies:

$$e^{W^T X} = 1 \quad (7)$$

$$W^T X = \sum_{i=0}^n \omega_i x_i = 0 \quad (8)$$

When training a model, the optimal values of the model parameters are determined so that  $h(x)$  correctly determines the class  $y$  for the input parameters  $x$ . The loss function  $L(h(x), y)$  defines the measure of the deviation of the hypothesis value from the exact value, on a single piece of data. The error function is the average of the loss function values on all data from the observed set:

$$J(\omega) = \frac{1}{m} \sum_{i=1}^m L(h(x^{(i)}), y^{(i)}) \quad (9)$$

In linear regression, the error function uses the mean squared deviation of  $h(x)$  from  $y$ ; however, due to the nature of the logistic classifier and the need for the error and loss functions to be convex, it is not a suitable loss function for logistic regression. The cross-entropy loss function meets the requirements for a logistic regression loss function, as follows:

$$L(h(x), y) = -y \ln h(x) - (1 - y) \ln 1 - h(x) \quad (10)$$

### 3.4.3. Support Vector Machine

Similar to logistic regression, with SVM, it is necessary to find a hyperplane that separates data belonging to different classes. Unlike LR, SVM has only a classification decision as an output, and it represents a non-probabilistic classifier.

If we look at the example of a binary classifier, and if the data can be linearly separated, this means that it is possible to construct two parallel hyperplanes that separate the data of different classes. The area of space between the classes is called the margin, and its value should be maximal so that the classification error of the new data is minimal. If  $n$  is the number of decisions used in the model, then the hypothesis is of the form:

$$h(x) = \omega_0 + \omega_1 x_1 + \omega_2 x_2 + \dots + \omega_n x_n = W^T X + \omega_0, \quad (11)$$

and the equation of the separating hyperplane is:

$$h(x) = W^T X + \omega_0 = 0. \quad (12)$$

The factors  $W$  and  $\omega_0$  can be chosen arbitrarily, but the convention is to choose a value such that the following applies to the support vectors  $X^{(sv)}$ :

$$y^{(sv)} \left( W^T X^{(sv)} + \omega_0 \right) = 1. \quad (13)$$

SVM has very good performance in a wide range of problems and much lower tendency to overfit than other methods. Unfortunately, the output is not of the probabilistic type, and depending on the number of features and the amount of data, it can be much slower than other models.

### 3.5. Multi-Class Classification

A naive Bayesian model is directly applicable to multiclass classification because nothing is assumed about the number of output values in the development of the model. Logistic regression and the support vector method can be applied to multiclass classification by combining the results of a number of binary classifiers in one of the following ways:

- One-vs-All (OvA) or One-vs-Rest (OvR) approach
- One-vs-One (OvO) approach

With the OvA approach,  $k$  binary classifiers for  $k$  classes are constructed. Each of the classifiers receives one class and treats all other classes together as another class. The problem with this principle is the imbalance of the number of examples in individual classifiers, as the number of examples in the second class is far greater than the number of examples in the first. The new data is classified into the class whose binary classifier produces the highest probability of the data belonging to the observed class.

For the OvO approach,  $\frac{k \cdot (k-1)}{2}$  binary classifiers are constructed—one for each pair of classes. The number of classifiers is much larger than in the OvA approach, but the training data set for each binary classifier is smaller. The new data is classified into the class selected by the most binary classifiers.

Multinomial logistic regression is a natural extension of logistic regression to work with multiple classes. The probability of belonging to a class is obtained using the function:

$$P(y = t|x) = e^{\sum_{i=0}^n \omega_i^{[t]} x_i} \sum_{j=1}^k e^{\sum_i^{[j]} x_i} = \frac{e^{(W^{[t]})^T x}}{\sum_{j=1}^k e^{(W^{[j]})^T x}}, \quad (14)$$

where  $k$  is the number of classes,  $n$  is the number of features,  $\omega_i^{[t]}$  is the weight parameter of the  $i$ th feature for the  $t$  class, and  $X$  is the feature value vector.

### 3.6. Assessment of Classifier Quality

When evaluating the classifier, a new data set that was not used for learning is used. Then, a pre-known class from the data set is compared with the class determined by the classifier. To compare and evaluate the performance of the classifiers, the following evaluation functions are used: accuracy, precision, recall, and f-measure.

In order to define functions for model evaluation, it is necessary to first explain the confusion matrix on the example of a binary classifier with the classes 0 and 1. Then, for each data record we want to classify, we distinguish four states: true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

The accuracy of the classifier represents the percentage of successfully classified data. The precision of the classifier represents the percentage of truly positive data. The recall of a classifier is a measure of the opposite of precision—of the data that are positive, what percentage is selected as positive. Combining precision and recall achieves an f1 measure.

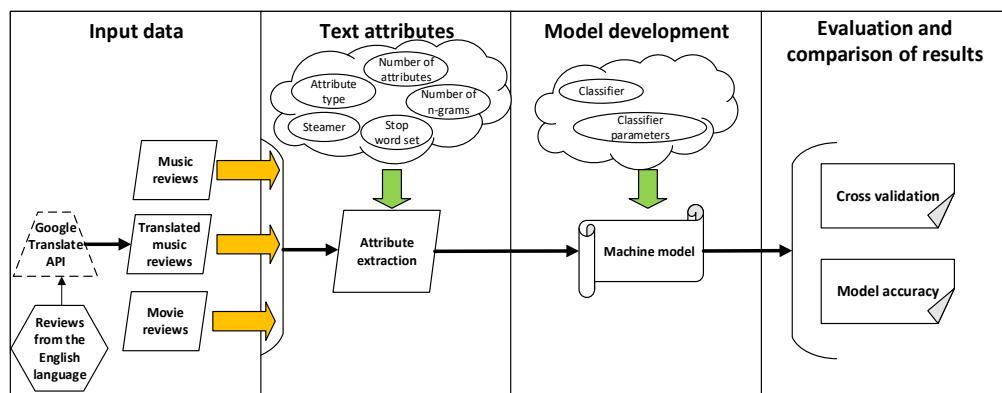
The technique of  $n$  cross-validations involves dividing the data set into  $n$  parts and then into  $n$  iterations— $n - 1$  parts are used for training the model and one part is used for validation. Finally, the intersection of all obtained values is taken as a result of the evaluation function.

#### 4. Results

The models were trained on each of the three described sets, while only the data sets of music reviews, written in Serbian, were used for testing. The parameters that were tested and adjusted in the development of the model can be divided into several groups:

- the model and input values of the machine learning model
- number of attributes and stop words
- number of n-grams
- value and attribute type

In this research, the existing stemmers for the Serbian language (Milošević [39]) and for the Croatian language (Ljubešić and Pandžić [40,41]) were used because they belong to a group of similar Slavic languages. The model evaluation diagram is shown in Figure 2. The optimal algorithm and parameters, as well as text attributes, were found by the method of examining different combinations with the help of the Pipeline and GridSearchCV classes from the scikit-learn library. Model accuracy was used as a function for model evaluation and comparison, as was the cross-validation technique, with  $n = 5$ .



**Figure 2.** Diagram of model evaluation.

##### 4.1. Results of the Three-Class Classification

First, three sentiment classes in the data set were considered. The results obtained when using a data set of music reviews for both training and testing the model are shown in Table 5 (Results (A)). Using the same parameter values, the model was trained on a set of movie reviews (Results (B) in Table 5), as well as music reviews translated from English (Results (C) in Table 5). The model testing was done with a set of music reviews originally written in Serbian.

**Table 5.** Results of the three-class classification.

	MNB	LR	SVM
Attribute type	bag of words	bag of words	TF-IDF
Attribute value	binary	binary	binary
Stemmer	Milošević	Milošević	Ljubešić/Pandžić
Number of n-gram	2	2	1
Max frequency n-gram	0.7	0.7	1
Min frequency n-gram	1	1	1
Number of attributes	20,000	20,000	20,000
Small letters only	yes	yes	yes
Results (A)	0.58	0.60	0.59
Results (B)	0.55	0.58	0.51
Results (C)	0.46	0.50	0.50

In the case of the three-class classification, approximately similar results were obtained when the model was trained on a set of movie reviews and tested on a set of music reviews.

The reason for this is that both data sets had reviews from the same portals, and so the review writing style and vocabulary were similar, even though they were different domains. In addition, expanding the data set improved the quality and precision of the developed model. When using the translated data set, the results were lower than the results obtained using only music reviews.

#### 4.2. Binary Classification Results

Only positive and negative reviews were observed from the input data set. The results obtained during the development of the model are shown in Table 6 (Results (A)). In this case, the same models were also trained on a set of film reviews, as well as music reviews translated from English, and they were tested on the original Serbian reviews (Results (B) and (C) in Table 6).

**Table 6.** Results of binary classifiers.

	MNB	LR	SVM
Attribute type	bag of words	bag of words	ag of words
Attribute value	binary	binary	binary
Stemmer	Ljubešić/Pandžić	Ljubešić/Pandžić	Ljubešić/Pandžić
Number of n-gram	1	1	3
Max frequency n-gram	1	0.7	1
Min frequency n-gram	1	1	1
Number of attributes	5000	5000	max
Small letters only	yes	yes	yes
Results (A)	0.77	0.75	0.77
Results (B)	0.72	0.61	0.73
Results (C)	0.62	0.45	0.60

With the binary classification, as with the three-class, we noticed that the results were very similar when training on a set of movie reviews. The results of using translated reviews were better in the three-class classification than they were in the binary classification.

#### 4.3. Hybrid Models

The last step in the model evaluation was the implementation and testing of a hybrid model: a naive Bayes–method of support vectors (SVM) hybrid [42]. This model was based on combining the linear model with the Bayesian model and replacing the word frequency attributes with their ratio vector of the NB counting of positive and negative classes. The main model was a linear classifier:

$$y_{(i)} = \text{sign}\left(W^T x_{(i)} + \epsilon\right) \quad (15)$$

If  $f^{(i)}$  is a vector of attributes and the output value is  $y_i$ ,  $V$  is the set of attributes and  $f_j^{(i)}$  is the number of occurrences of the attribute  $V_j$  in the input text  $i$ .

The counting vectors of the positive and negative class are defined as:

$$p = \alpha + \sum_{i:y^{(i)}=1} f^{(i)} \quad (16)$$

$$q = \alpha + \sum_{i:y^{(i)}=-1} f^{(i)} \quad (17)$$

The positive to negative class count ratio vector is defined as:

$$r = \log\left(\frac{p/\|p\|_1}{q/\|q\|_1}\right) \quad (18)$$

In order to combine the above equations, an elemental multiplication of the SVM vector of attributes ( $f$ ) and the ratio vector of the results of NB counting the positive and negative classes of ( $r$ ) is performed:

$$\bar{f}^{(k)} = r \cdot f^{(k)} \quad (19)$$

The resulting vector is used as the input for a standard SVM classifier.

The evaluation results of the described model are shown in Table 7. The model was trained on different data sets, while testing was always done on the data set of music reviews.

**Table 7.** NB–SVM hybrid model results.

	Three Classes	Binary
Results (A)	0.57	0.78
Results (B)	0.54	0.70
Results (C)	0.49	0.62

The research tested the use of logistic regression models instead of SVM. The results obtained by combining the naive Bayes and logistic regression are shown in Table 8.

**Table 8.** NB–LR hybrid model results.

	Three Classes	Binary
Results (A)	0.58	0.79
Results (B)	0.51	0.74
Results (C)	0.42	0.61

## 5. Discussion

In the previous sections, we described the application of the traditional machine learning techniques to the problem of multilingual sentiment analysis in NLP. We used the Serbian data set in our experimental set-up.

In addition to standard algorithms, such as LR, SVM and MNB, the hybrid algorithms NB–SVM and NB–LR were considered due to the problem of binary classification. As suggested in [42], in LR, SVM, and NB–SVM, the  $L_2$  loss function and  $L_2$  regularization were used. A five-layer nested stratified cross-validation was used for the optimization of hyperparameter  $C$ , which is used in LR, SVM, and NB–SVM algorithms, as well as for the optimization of hyperparameter  $\beta$  in the NB–SVM algorithm. All other model hyperparameters were set to default values. During classification, all text was normalized to lowercase letters.

Two different types of stemmers were used in the three-class classification. In Table 6, we can see that in the case of binary classification, the Ljubešić and Pandžić stemmer was used as the optimal solution for morphological normalization.

Overall, from the results shown in Tables 5–8, it can be seen that the hybrid approach in the form of the naive Bayesian model and the linear classifier offers average good results, but it still does not provide significant improvements compared to other models. A 2% improvement can be seen with the binary classifiers as there is a clear separation in positive and negative sentiment. In the three-class classifiers, the neutral class is not clearly separated from the other two. However, it represents a combination of positive and negative sentiments in the review, and so the hybrid model and approach using the ratio vectors of the NB class counts does not contribute to the quality of the model. Correction of typographical errors, normalization of emoticons, and character repetitions and morphological normalization are useful for all sentiment analysis problems when applied with features obtained by the bag-of-words principle.

Sentiment annotation was performed, and data sets were realized in the Serbian language using 13 different sources in Serbian (web portals) and MARD, originally written in English [38], with two approaches: the original Serbian language and a machine translation of the content, from English to Serbian, using Google Translate. In this way, the collected data can help other researchers to improve the machine translation process into the Serbian language. Furthermore, concerning the research papers discussed in Section 2, the accuracy obtained in this research is in the range of other results of multilingual models for the Serbian language.

## 6. Conclusions

This research aimed to collect data in a low-resource language and develop a model for sentiment analysis in the Serbian language. In the most extensive and state-of-the-art research in multilingual sentiment analysis, languages with limited resources, such as Serbian, are not covered, or they are only covered to a small extent [20,21].

In addition to the movie reviews collected in [37], a data set of music reviews (originally written in the Serbian language) is another applied set used for sentiment analysis. With the increase in the set and the scope of the data, the opportunities for developing new models and sentiment analysis in the Serbian language also increase. Likewise, the research showed that a set of movie and music reviews can be used together and that the models developed in this case offered good results. The assumption is that one of the reasons for the good results is the fact that part of the data from both data sets was collected from the same or similar portals.

The problem of the unavailability of resources in the Serbian language was attempted to be overcome by using an English data set which was translated into the Serbian language using the Google Translate API. Other researchers have also used Google Translate or the Bing translator to work on translated data for multilingual or cross-lingual sentiment analysis [43–49]. However, we did not achieve good results in this research, and the model had a much weaker performance than when working with reviews originally written in Serbian, likely because of the different vocabulary and style of writing reviews, as well as the quality of the translated text.

The results of this research represent a breakthrough in developing machine processing in the Serbian language. Furthermore, the creation of available annotated data sets with reviews will facilitate the further development of the sentiment analysis of short texts in the Serbian language. The main contributions of this research are the creation of a representative and sufficiently large database with movie and music reviews in the Serbian language from various sources available on the Internet, the application of the most significant algorithms in supervised text classification, and the development of different models that were trained on a set of collected data, after which the evaluation was carried out.

Using additional and/or more advanced techniques for extracting attributes from text, the proposed models can be further improved. In this paper, negation was not processed, and the filtering of stop words was done automatically within the existing library implementation of the algorithm for creating attribute vectors. A more detailed analysis of the vocabulary in the data sets could create a better set of stop words, followed by testing them on the given models. These are also the main limitations of the study. In the continuation of the research, the authors will also replace the traditional machine learning methods with a CNN or LSTM in order to obtain even better precision with more modern models, while still requiring minimal execution time.

**Author Contributions:** Conceptualization, D.D. and D.Z.; methodology, B.N.; related work, D.D.; software, D.Z. and D.D.; validation, D.Z.; writing—original draft preparation, D.D. and B.N.; visualization, D.D.; supervision, D.D. and B.N.; project administration, B.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Science Fund of the Republic of Serbia, grant no. 6526093, AI-AVANTES (<http://fondznanuku.gov.rs/>).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## Abbreviations

The following abbreviations are used in this manuscript:

API	Application programming interface
IDF	Inverse document frequency
LR	Logistic regression
MARD	Multimodal album reviews data set
ML	Machine learning
MNB	Multinomial naïve Bayes
NB	Naïve Bayes
NLP	Natural language processing
OvA	One-vs-All
OvO	One-vs-One
SVM	Support vector machine
TF	Term frequency

## References

- Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs Up? Sentiment Classification using Machine Learning Techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), EMNLP, Philadelphia, PA, USA, 6–7 July 2002.
- Abbasi, A.; Chen, H.; Salem, A. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Trans. Inf. Syst.* **2008**, *26*, 1–34. [[CrossRef](#)]
- Das, S.R.; Chen, M.Y. Yahoo! for Amazon: Sentiment extraction from small talk on the Web. *Manag. Sci.* **2007**, *53*, 1375–1388. [[CrossRef](#)]
- Neethu, M.S.; Rajasree, R. Sentiment analysis in Twitter using machine learning techniques. In Proceedings of the 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, India, 4–6 July 2013.
- Bouazizi, M.; Ohtsuki, T. Sentiment analysis: From binary to multi-class classification: A pattern-based approach for multi-class sentiment analysis in Twitter. In Proceedings of the IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 22–27 May 2016.
- Čutura, G.; Knežević, B.; Drašković, D. Public opinion about Novak Djokovic through the eyes of Twitter. In Proceedings of the 12th International Conference on Information Society and Technology, Kopaonik, Serbia, 13–16 March 2022; pp. 81–85.
- Benjamin, M. Hard Numbers: Language Exclusion in Computational Linguistics and Natural Language Processing. In Proceedings of the LREC 2018 Workshop “CCURL2018–Sustaining Knowledge Diversity in the Digital Age”, Miyazaki, Japan, 7–12 May 2018; pp. 26–32.
- El-Haj, M.; Kruschwitz, U.; Fox, C. Creating language resources for under-resourced languages: Methodologies, and experiments with Arabic. *Lang. Resour. Eval.* **2015**, *49*, 549–580. [[CrossRef](#)]
- Maxwell, M.; Hughes, B. Frontiers in linguistic annotation for lower-density languages. In Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora. Association for Computational Linguistics, Sydney, NSW, Australia, 22 July 2006; pp. 29–37.
- Streiter, O.; Scannell, K.P.; Stuflesser, M. Implementing NLP projects for noncentral languages: Instructions for funding bodies, strategies for developers. *Mach. Transl.* **2006**, *20*, 267–289. [[CrossRef](#)]
- Towards Data Science. Available online: <http://towardsdatascience.com/major-trends-in-nlp-a-review-of-20-years-of-acl-research-56f5520d473> (accessed on 15 May 2022).
- Kornai, A. Digital Language Death. *PLoS ONE* **2013**, *8*, e77056.
- Berment, V. Several directions for minority languages computerization. In Proceedings of the 19th International Conference on Computational Linguistics: Project Notes (COLING 2002). Association for Computational Linguistics, Taipei, Taiwan, 26–30 August 2002.
- King, B.P. *Practical Natural Language Processing for Low-Resource Languages*; University of Michigan: Ann Arbor, MI, USA, 2015.
- Duong, L.T. Natural Language Processing for Resource-Poor Languages. Ph.D. Thesis, University of Melbourne, Melbourne, VIC, Australia, 2017.
- Pang, B.; Lee, L. Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.* **2008**, *2*, 1–135. [[CrossRef](#)]

17. Liu, B.; Zhang, L. A Survey of Opinion Mining and Sentiment Analysis. In *Mining Text Data*; Aggarwal, C.C., Zhai, C., Eds.; Springer: Boston, MA, USA, 2012; pp. 415–463.
18. Paulino, J.; Almirol, L.; Favila, J.; Aquino, K.; De La Cruz, A.; Roxas, R. Multilingual Sentiment Analysis on Short Text Document Using Semi-Supervised Machine Learning. In Proceedings of the 5th International Conference on E-Society, E-Education and E-Technology, Virtual Format, 21–23 August 2021; pp. 164–170.
19. Nankani, H.; Dutta, H.; Shrivastava, H.; Rama Krishna, P.V.N.S.; Mahata, D.; Shah, R.R. Multilingual Sentiment Analysis. In *Deep Learning-Based Approaches for Sentiment Analysis*; Part of the Algorithms for Intelligent Systems Book Series; Agarwal, B., Nayak, R., Mittal, N., Patnaik, S., Eds.; Springer: Singapore, 2020.
20. Dashtipour, K.; Poria, S.; Hussain, A.; Cambria, E.; Hawalah, A.Y.; Gelbukh, A.; Zhou, Q. Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques. *Cogn. Comput.* **2016**, *8*, 757–771. [[CrossRef](#)]
21. Sagnika, S.; Pattanaik, A.; Mishra, B.S.P.; Meher, S. A Review on Multi-Lingual Sentiment Analysis by Machine Learning Methods. *J. Eng. Sci. Technol. Rev.* **2020**, *13*, 154–166. [[CrossRef](#)]
22. Bera, A.; Ghose, M.K.; Pal, D.K. Sentiment Analysis of Multilingual Tweets Based on Natural Language Processing (NLP). *Int. J. Syst. Dyn. Appl.* **2021**, *10*, 1–12. [[CrossRef](#)]
23. Xu, H.; Van Durme, B.; Murray, K. BERT, mBERT or BiBERT? A Study on Contextualized Embeddings for Neural Machine Translation. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Online, 7–11 November 2021.
24. Khan, L.; Amjad, A.; Ashraf, N.; Chang, H.-T. Multi-class sentiment analysis of urdu text using multilingual BERT. *Sci. Rep.* **2022**, *12*, 5436. [[CrossRef](#)] [[PubMed](#)]
25. Pota, M.; Ventura, M.; Fujita, H.; Esposito, M. Multilingual evaluation of pre-processing for BERT-based sentiment analysis of tweets. *Expert Syst. Appl.* **2021**, *181*, 115119. [[CrossRef](#)]
26. Agüero-Torales, M.; Salas, J.; López-Herrera, A. Deep learning and multilingual sentiment analysis on social media data: An overview. *Appl. Soft Comput.* **2021**, *107*, 107373. [[CrossRef](#)]
27. Kanfoud, M.R.; Bouramoul, A. SentiCode: A new paradigm for one-time training and global prediction in multilingual sentiment analysis. *J. Intell. Inf. Syst.* **2022**; *Online ahead of print*. [[CrossRef](#)] [[PubMed](#)]
28. Žitnik, S.; Blagus, N.; Bajec, M. Target-level sentiment analysis for news articles. *Knowl.-Based Syst.* **2022**, *249*. [[CrossRef](#)]
29. Ljubešić, N.; Lauc, D. BERTić-The transformer language model for Bosnian, Croatian, Montenegrin and Serbian. *arXiv* **2021**, arXiv:2104.09243, 2021.
30. Mozetić, I.; Grčar, M.; Smailović, J. Multilingual Twitter Sentiment Classification: The Role of Human Annotators. *PLoS ONE* **2016**, *11*, e0155036. [[CrossRef](#)]
31. Ljajić, A.; Marovac, U. Improving Sentiment Analysis for Twitter Data by Handling Negation Rules in the Serbian Language. *Comput. Sci. Inf. Syst.* **2018**, *16*, 289–311. [[CrossRef](#)]
32. Batanović, V. Semantic Similarity and Sentiment Analysis of Short Texts in Serbian. In Proceedings of the 29th Telecommunications Forum (TELFOR), Virtual Event, 11 December 2021.
33. Lohar, P.; Popovic, M.; Way, A. Building English-to-Serbian Machine Translation System for IMDb Movie Reviews. In Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing, Florence, Italy, 2 August 2019; pp. 105–113.
34. Kumpulainen, I.; Praks, E.; Korhonen, T.; Ni, A.; Rissanen, V.; Vankka, J. Predicting Eurovision Song Contest Results Using Sentiment Analysis. In *Artificial Intelligence and Natural Language*; Filchenkov, A., Kauttonen, J., Pivovarova, L., Eds.; Springer International Publishing: Cham, Switzerland, 2020; Volume 1292.
35. Mladenović, M.; Mitrović, J.; Krstev, C.; Vitas, D. Hybrid sentiment analysis framework for a morphologically rich language. *J. Intell. Inf. Syst.* **2016**, *46*, 599–620. [[CrossRef](#)]
36. Stankovic, R.; Kosprdic, M.; Ikonic-Nesic, M.; Radovic, T. Sentiment Analysis of Sentences from Serbian ELTeC corpus. In Proceedings of the SALLD-2 Workshop at Language Resources and Evaluation Conference (LREC), Marseille, France, 24 June 2022; pp. 31–38.
37. Batanovic, V.; Nikolic, B.; Milosavljevic, M. Reliable Baselines for Sentiment Analysis in Resource-Limited Languages: The Serbian Movie Review Dataset. In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), LREC, Portorož, Slovenia, 23–28 May 2016.
38. Oramas, S.; Espinosa-Anke, L.; Lawlor, A.; Serra, X.; Saggion, H. Exploring Customer Reviews for Music Genre Classification and Evolutionary studies. In Proceedings of the 17th International Society for Music Information Retrieval Conference, New York, NY, USA, 7–11 August 2016.
39. Milošević, N. Stemmer for Serbian language. *arXiv* **2012**, arXiv:1209.4471.
40. Ljubešić, N.; Boras, D.; Kubelka, D. Retrieving Information in Croatian: Building a Simple and Efficient Rule-Based Stemmer. In Proceedings of the 1st International Conference The Future of Information Sciences—INFUTURE: “Digital Information and Heritage”, Zagreb, Croatia, 7–9 November 2007.
41. Ljubešić, N.; Klubička, F.; Agić, Ž.; Jazbec, I.-P. New Inflectional Lexicons and Training Corpora for Improved Morphosyntactic Annotation of Croatian and Serbian. In Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016). European Language Resources Association (ELRA), Portorož, Slovenia, 23–28 May 2016; pp. 4264–4270.
42. Wang, S.; Manning, C.D. Baselines and Bigrams: Simple, Good Sentiment and Topic Classification. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012), Jeju Island, Korea, 8–14 July 2012; pp. 90–94.

43. Hogenboom, A.; Heerschap, B.; Frasincar, F.; Kaymak, U.; de Jong, F. Multi-lingual support for lexicon-based sentiment analysis guided by semantics. *Decis. Support Syst.* **2014**, *62*, 43–53. [[CrossRef](#)]
44. Lin, Z.; Jin, X.; Xu, X.; Wang, Y.; Tan, S.; Cheng, X. Make it possible: Multilingual sentiment analysis without much prior knowledge. In Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), IEEE Computer Society, Warsaw, Poland, 11–14 August 2014; Volume 2, pp. 79–86.
45. Hajmohammadi, M.S.; Ibrahim, R.; Selamat, A.; Fujita, H. Combination of active learning and self-training for crosslingual sentiment classification with density analysis of unlabelled samples. *Inf. Sci.* **2015**, *317*, 67–77. [[CrossRef](#)]
46. Becker, K.; Moreira, V.P.; dos Santos, A.G. Multilingual emotion classification using supervised learning: Comparative experiments. *Inf. Processing Manag.* **2017**, *53*, 684–704. [[CrossRef](#)]
47. Chen, Z.; Shen, S.; Hu, Z.; Lu, X.; Mei, Q.; Liu, X. Ermes: Emoji-Powered Representation Learning for Cross-Lingual Sentiment Classification. *arXiv* **2018**, arXiv:1806.02557.
48. Balahur, A.; Turchi, M. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. *Comput. Speech Lang.* **2014**, *28*, 56–75. [[CrossRef](#)]
49. Bhargava, R.; Sharma, Y. MSATS: Multilingual sentiment analysis via text summarization. In Proceedings of the 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, IEEE, Noida, India, 12–13 January 2017; pp. 71–76.