DOI:10.25092/baunfbed. 1596321 *J. BAUN Inst. Sci. Technol., 27(2), 569-586, (2025)*

Classification of news about Turkey in the foreign press through text mining

Murat ISIK[1,*], Emrah AYDEMİR[2]

[1] *Department of Computer Engineering, Kirsehir Ahi Evran University, Kirsehir, TURKEY,*
[2] *Management Information Systems, Sakarya University, Sakarya, TURKEY.*

Geliş Tarihi (Received Date): 04.12.2024
Kabul Tarihi (Accepted Date): 03.04.2025

Abstract

In recent years, many newspapers and news providers have begun presenting their content via web pages or through social media. This shift has led to a massive increase in the volume of news content available, necessitating the analysis and management of this vast information flow. In this study, 8,385 pieces of news content related to Turkey were collected from the web pages of major foreign news content providers, including Fox, The Guardian, BBC, and CNN. While traditional techniques classify news texts into categories based on their content, this study achieved an average accuracy rate of 89.36% by classifying the contents according to eight predefined areas of interest. Moreover, analyses were conducted based on the publication dates of all foreign news content, revealing relationships between the dates of publication and the classified areas of interest. Additionally, sentiment analysis was conducted on the collected foreign news content using the BERT algorithm, which identified the sentiment categories of the contents and examined the perception of Turkey in the foreign press.

Keywords: *Sentiment analysis, BERT, text mining, data mining, news analysis, classification of news, relationship between news dates and content.*

Metin madenciliği ile Türkiye'nin yabancı basındaki imajının analizi

Öz

Son yıllarda, birçok gazete ve haber sağlayıcısı içeriklerini web sayfaları ve sosyal medya aracılığıyla sunmaya başlamıştır. Bu dönüşüm, haber içeriklerinin hacminde büyük bir

[*]Murat IŞIK, muratisik@ahievran.edu.tr, https://orcid.org/0000-0003-3200-1609
Emrah Aydemir, aydemir.emrah23@gmail.com, https://orcid.org/0000-0002-8380-7891

artışa yol açarak bu geniş bilgi akışının analiz edilmesi ve yönetilmesini zorunlu hale getirmiştir. Bu çalışma kapsamında, Fox, The Guardian, BBC ve CNN gibi büyük yabancı haber sağlayıcılarının web sayfalarından Türkiye ile ilgili 8.385 haber içeriği toplanmıştır. Geleneksel teknikler, haber metinlerini içeriklerine göre sınıflandırırken, bu çalışmada içerikler önceden tanımlanmış ilgi alanlarına göre sınıflandırılarak %89,36'lık ortalama bir doğruluk oranı elde edilmiştir. Ayrıca, yabancı haber içeriklerinin yayımlanma tarihleri temelinde yapılan analizler, yayımlanma tarihleri ile sınıflandırılmış ilgi alanları arasındaki ilişkiler ortaya çıkarılmıştır. Bunun yanı sıra, BERT algoritması kullanılarak toplanan haber içeriklerinde duygu analizi gerçekleştirilmiş, içeriklerin duygu kategorileri belirlenmiş ve Türkiye'nin yabancı basındaki algısı incelenmiştir.

Anahtar kelimeler: *Duygu analizi, BERT, metin madenciliği, veri madenciliği, haber analizi, haber sınıflandırması, haber tarihleri ve içerik ilişkisi.*

1. Introduction

Since the mid-1980s, there has been an unprecedented explosion in the capacity to generate, store, and transmit data worldwide, especially in digital format [1]. This data explosion has largely produced textual content, which is often found in an unstructured format. Being unstructured implies that the textual data is free-form, containing ambiguities in words and sentences [2]. The unstructured nature of textual data has necessitated the development of specialized tools for its storage, processing, and extraction of meaningful information. At this juncture, Text Mining techniques have emerged as a powerful tool in the analysis of large textual datasets and the extraction of meaningful information from them [3]. Text mining, gaining popularity with large data sources [3, 4], is the process of extracting previously unknown information from large volumes of unstructured text [5, 6].

In the field of text mining, a significant focus has been placed on the classification of news content [2]. This emphasis stems from the fact that news content is among the most crucial textual materials, impacting various segments of society and desired to be classified [4]. Indeed, news has transcended merely announcing current events; it now shapes societal thought structures, influences values, and fosters participation in democratic processes. In this context, the transformation of news content from merely a source of information to a social force that enhances awareness of societal events has been impactful. Today, accessing news content has never been easier, faster, or more effortless. Thanks to online news providers [7] and social media platforms [8, 9], news is readily accessible. However, this ease of access has led to a massive increase in the volume of news available, complicating users' ability to reach news of interest [10] and making the analysis of news content increasingly challenging.

The primary objective of this study is to collect and analyze news related to Turkey in the foreign press. This analysis aims to examine the relationships between news contents about Turkey in the foreign media and to conduct sentiment analyses. This goal will be achieved through two main objectives. The first objective is to successfully classify news contents according to the areas of interest of the user. The second objective is to analyze and resolve foreign news content based on their publication dates, classification results, and the sentiment categories they possess.

Unlike previous studies that classify news based on categories predefined by content providers, this study introduces a novel approach by categorizing news according to eight areas of interest and performing sentiment analysis using BERT. The integration of traditional machine learning models (RF, KNN, SVM, MNB) with BERT's deep learning capabilities enhances the robustness of the analysis.

The paper is structured as follows: The introduction outlines the research gap and objectives. The related work section discusses previous studies on text mining and sentiment analysis. The methods section details data collection, preprocessing, and model training techniques. The results section presents performance metrics for each model, and the discussion evaluates the implications of these results. Finally, the conclusion summarizes key findings and suggests areas for future research.

2. Related literature

Hassani and colleagues [3] emphasized the power of text mining within big data analytics in their study and explored its potential applications. Furthermore, they offered suggestions for successful analyses of news content using text mining techniques. In this context, it is evident that text mining is highly effective in analyzing news content. Some studies concerning the classification of news content and the use of the BERT algorithm in these contents are presented below.

Leonard and colleagues [11] have utilized news headlines to classify news content into existing categories. They achieved 90% accuracy with their SVC model-based system and conducted comparisons with Linear Regression, Multinomial Naive Bayes, Decision Tree, and Random Forest models. The study utilized news content collected from a single source. It was used for classifying news into pre-defined categories rather than according to areas of interest. Moreover, no analysis of the relationships between news publication dates or sentiment analysis was conducted.

Agarwal and colleagues [12] used Random Forest, Decision Tree, k-Nearest Neighbors Classifier, and Gaussian Naive Bayes algorithms to classify news according to areas of interest, achieving the highest result with Random Forest at 91.94%. The study primarily classified news contents based on categories determined by the content provider, rather than by areas of interest. The news contents were collected from a single source. The study did not focus on examining areas of interest, conducting sentiment analysis, or analyzing the historical relationships of the news.

Sadjadi and colleagues [13] proposed a semi-supervised model that works with both labeled and unlabeled data for classifying news content across two levels. In their study, they examined data sets collected from two different sources, classifying the first data set into four categories and the second into five categories, according to categories determined by the content providers. They achieved an accuracy rate of 77.3% for the first data set and 81.92% for the second. The classification was carried out according to the categories defined by the news providers, and different numbers of categories were used in different data sets. Unlike the traditional category-based classification, the proposed study conducted classification based on areas of interest, utilizing eight different interest areas across content obtained from all providers. The success rate of the proposed approach was relatively higher compared to the results of the study examined.

In a 2018 study, Miao and colleagues [14] employed the TF-IDF feature extractor and worked on classification models based on KNN, NB, and SVM algorithms. They achieved the highest F1-Score with the SVM algorithm at 95.7%. In 2021, Salehin and associates [15] collected news in the Bangla language and classified it according to categories defined by the publisher using various machine learning models (LR, MNB, SVM, RF, XGBoost, MLP, and LSTM). They opted for TF-IDF as the feature extractor and worked on a total of 75,951 news contents. The highest accuracy was obtained with the LSTM model at 87%. However, these studies conducted classification based on categories predefined by the content providers and worked with content obtained from a single source. Additionally, they employed only one type of feature extractor. No sentiment analysis or analysis of news publication dates was conducted in these studies.

Lin and colleagues [16] conducted a study where they utilized BERT to analyze news content, enabling the detection of harmful news. Their proposed method achieved an F1-score of 66.3%. Additionally, the developed method demonstrated good performance in identifying cases of information distortion. In this reviewed study, data indicating the emotional state of news content was used to discern the relationship between harmful and fake news. In the proposed study, however, BERT were utilized in evaluating the overall condition of the news.

Hayawi and colleagues [17] developed a model that employs XGBoost, LSTM, and BERT to perform sentiment analysis on news about the COVID-19 vaccine, achieving an accuracy rate of 98% with BERT. Similarly, To and colleagues [18] developed a new model to identify anti-vaccine sentiments during the COVID-19 vaccination process, achieving the best result with a BERT-based SVM algorithm at 92.3%. In these studies, the value derived from BERT was used solely for classification purposes. In the proposed study, however, it has been used for the analysis of news content.

Ahmet et al. [19] reviewed several existing approaches for classifying online news articles and discussed a framework for the automatic classification of these articles. They achieved a 93% accuracy rate using a Bayesian classifier among methods including NB, KNN, SVM, and LR. Their analysis was limited to categories previously determined by the news content provider. Additionally, their study did not encompass emotion analysis or investigate the publication dates of the news.

In the literature, there are numerous studies related to the classification of news articles [20-23] ; however, it has been observed that the use of the BERT algorithm in news content generally focuses on the detection of fake news [24-26]. Most existing studies focus on classifying news into categories predefined by content providers, which limits the scope of analysis to surface-level understanding. This study addresses this gap by leveraging BERT for sentiment analysis in combination with traditional classifiers to provide a dual-layer analysis. The application of BERT in this study is innovative as it enables the detection of nuanced sentiment across different areas of interest, providing deeper insights into how Turkey is portrayed in the foreign press. Unlike previous works that used BERT solely for classification, this study employs it to reveal sentiment trends over time and across different news categories. Additionally, no studies were found that analyze the relationship between the publication dates of news articles.

3. Methods

In the transformation of unstructured news texts into numerical data, feature extractors such as TF-IDF (Term Frequency-Inverse Document Frequency) [27], WIDF (Weighted Inverse Document Frequency) [28], and GWS (Glasgow Weighting) [29] have been utilized. For the classification of news texts, classifiers such as RF (Random Forest) [30], SVM (Support Vector Machine) [31], KNN (K-Nearest Neighbor) [32], and MNB (Multinomial Naive Bayes) [33] have been employed. For the sentiment analysis of news content, the BERT (Bidirectional Encoder Representations from Transformers) algorithm [34], one of today's most popular sentiment analysis models, has been used.

During the implementation phase of the study, a computer equipped with an Intel (R) Core (TM) i7-10750H CPU (2.6 GHz), 32.00 GB RAM, a 64-bit Operating System, and an 8GB Nvidia GPU was utilized. In the development phase of the application, the Python programming language was used, incorporating libraries such as NLTK (Natural Language Toolkit) and Spacy. The architecture of the proposed model can be seen in Figure 1.

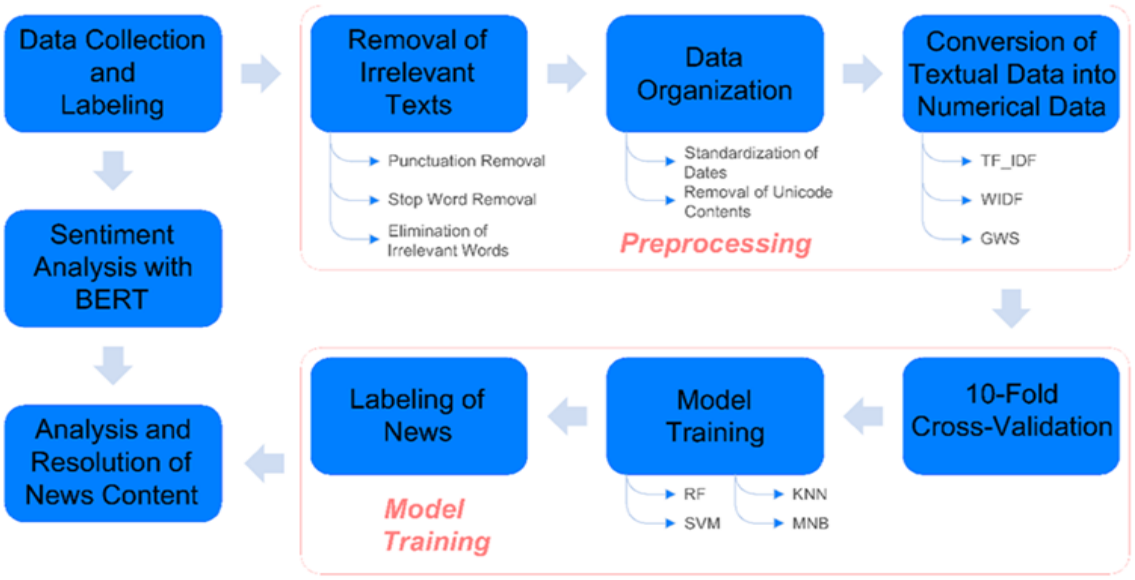

Figure 1. Model architecture.

3.1. Data collection and labelling

News content was collected from the web pages of news providers such as Fox, The Guardian, BBC, and CNN using an algorithm developed with BeautifulSoup and Selenium tools. The numbers and yearly distributions of the collected news contents are presented in Table 1. A sample from the dataset consisting of news articles spanning 2009 to 2023 is shown in Table 2. Since no news content related to Turkey was found from the news content providers prior to 2009, the news contents have been collected starting from the year 2009. Data not relevant to this study, such as news headlines, author name, have been included in the dataset to enable researchers to conduct various other studies. The dataset shared online for other researchers' use [35].

Table 1. Numbers of news collected by year.

Year	The Guardian	BBC	FOX	CNN
2009 – 2011	27	5	0	208
2012	4	3	0	525
2013	11	7	0	501
2014	10	14	11	420
2015	9	36	69	728
2016	31	47	116	830
2017	15	27	114	693
2018	16	13	121	719
2019	23	9	76	768
2020	19	13	59	402
2021	27	13	34	398
2022	54	23	85	484
2023	147	96	80	245
Total	393	306	765	6921

A subset of 750 news articles was manually labelled to ensure high-quality training data for the classification models. During this process, eight areas of interest—Erdogan, Refugees and Minorities, International Relations, Sports and Tourism, Elections, Natural Disasters and Epidemics, Terror, and Economy—were identified based on an empirical review of the labelled articles. Unlike studies that rely on predefined threshold values for classification, our approach was purely human-driven, with no automatic thresholding mechanisms. Each article was carefully reviewed and assigned a binary label (relevant or not relevant) for each area of interest according to standardized labeling criteria, which were consistently applied to minimize potential bias. This human-in-the-loop approach ensured that the predefined areas of interest accurately captured the main subjects covered in the news about Turkey, reflecting nuanced contextual relevance rather than frequency-based cutoffs. The manually labelled subset also served as a benchmark to evaluate and refine the performance of the automated classification models. Since some news items can cover more than one area of interest, these 2242 labels encompass the identified interests. In each category, half of the labels indicate relevance to the specific area, while the other half denote irrelevance. For example, in the 'International Relations' area, 302 news contents were labelled as 1 (relevant), and an equal number were labelled as 0 (irrelevant), totaling 604 labels. Table 3 shows the number of news articles according to the other labelled areas of interest.

Table 2. A sample from the dataset.

Headline	Content	Author	Date – Time	Link
Headline	Content	Author	Date – Time	Link
Biden arrives in Greece at piv…	Biden will meet with top Greek…	CNN Wire Staff	11:39 AM EST, Sun December 4, 2011	https://www.cnn...
Biden begins 4-day visit to Tur…	Biden arrives Thursday in …	CNN Wire Staff	5:39 PM EST, Thu December 1, 2011	https://www.cnn...
Turkey seeks to detain 70 offic…	ANKARA Turkey – Turkey's sta…	Associated Press	January 4, 2018 2:07am EST	https://www.fox...
Human rights row over BP…	BP and a consortium…	Paul Brown	Sat 31 Aug 2002 01.40 BST	https://www.theg...
Quiet end to Turkey's coll…	Every morning Yasemin …	Jonathan Head	31 December 2010	https://www.bbc...

Table 3. Number of labelled news articles.

Area of Interest	Count of Labeled News Articles
Erdogan	264
Refugees and Minorities	214
International Relations	604
Sports and Tourism	244
Elections	214
Natural Disaster and Epidemic	236
Terror	292
Economy	174

3.2. Preprocessing

Prior to model training, it is essential for news content to undergo several preprocessing steps. It is imperative to clean the text of content that serves no purpose for classification and to organize the data appropriately for its intended use. The primary goal at this stage is to focus on the words that can better convey the main ideas and concepts within the text. The preprocessing process has been carried out through the systematic steps presented each paragraph below.

In the news content, elements that do not contribute to the identification of the area of interest, such as author names, place names, content provider information, and details about photographs in the text (name of the photographer, location, description, and date information), have been removed. Additionally, words identified in some news contents that are not Unicode have also been eliminated. A simple algorithm has been developed to facilitate the removal of these contents.

In the text, content known only as function words, which do not convey any meaningful information and provide no valuable insight into the area of interest of the news text, has been removed. The removal of these contents utilized the default "Stop Words" removal feature of the Spacy library.

Upon examining Table 2, it is apparent that each content provider uses a different time format. To establish a temporal relationship between the news articles, these time formats have been standardized into a single format using a developed algorithm.

The presence of words that appear too frequently or infrequently in news content can affect the accuracy of classification results. For this reason, words that occur below a certain count and above a certain percentage have been removed during model training. However, since the textual content in each area of interest varies significantly, different filtering criteria have been used for each area. The criteria employed are presented in the Filtering column of Table 5.

During the digitization of textual content, feature extraction algorithms such as TF-IDF (Term Frequency-Inverse Document Frequency), WIDF (Weighted Inverse Document Frequency), and GWS (Glasgow Weighting Scheme) were utilized. The results obtained were analyzed to determine the most effective method for the dataset in use, and this method was then employed in the training of the final model.

3.3. Model training

For model training, classifiers such as RF (Random Forest), SVM (Support Vector Machine), KNN (K-Nearest Neighbors), and MNB (Multinomial Naive Bayes) were utilized. The separation of training and test data employed the technique of 10-fold cross-validation. Model performance was measured using a confusion matrix from which accuracy, sensitivity, precision, and F1-Score metrics were calculated.

3.4. Analysis of news content

In this step, sentiment analysis of news content is conducted using the BERT algorithm before preprocessing, and the relationships between the labeled news contents and their publication dates are visually depicted. The goal is to measure whether the news content possesses a positive or negative sentiment using BERT, and sentiment scores are recorded accordingly. The BERT model used in this study is the pre-trained BERT-base uncased model. The model was applied without custom fine-tuning, using its default architecture and parameters. The sentiment analysis was conducted using the default BERT tokenizer and model weights to classify news sentiment as positive or negative. Since we used the pre-trained BERT-base uncased model without fine-tuning, we did not train BERT on our dataset. Therefore, calculating F1-score for BERT is not applicable in this study, as it is primarily used for sentiment classification without custom optimization. Instead, performance metrics (accuracy, precision, recall, and F1-score) were reported for the classification models that were trained on our dataset.

When using the BERT algorithm, negative and positive scores typically represent the emotional tone or semantically oriented assessment of a text. These scores are part of the transformer-based language model used to understand relationships between texts or the contexts of specific words. For instance, in text analysis, a positive score given by BERT indicates that the text generally contains positive emotions or is used in a positive context, while a negative score suggests that the text carries negative or undesirable emotions or is used in a negative context.

3.5. Performance measures

In the evaluation of classification models, the confusion matrix is utilized as a fundamental tool, encapsulating the complete performance of the model across different classes [36]. This matrix categorizes predictions into true positives, true negatives, false positives, and false negatives, thus enabling a nuanced analysis of both successes and errors in classification. An example of a confusion matrix is presented in Table 4.

Table 4. Confusion matrix.

		Actual Classes	
		Positive	Negative
Predicted Classes	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

In Table 4, true positives (TP) occur when the model correctly predicts positive instances, while true negatives (TN) are cases where negative instances are correctly identified as such. Conversely, false positives (FP) arise when the model incorrectly predicts positive

outcomes for negative instances. False negatives (FN) occur when the model fails to identify an actual positive instance.

In assessing the performance of classification models, several key metrics are typically employed: accuracy, precision, recall, and the F1-Score, all derived from the confusion matrix. Accuracy, as presented in Equation 1, is defined as the ratio of true predictions to the total number of cases evaluated, providing an overall measure of effectiveness [36]. Precision (positive predictive value), shown in Equation 2, and recall (sensitivity), shown in Equation 3, assess the accuracy of positive predictions and the rate at which positive instances are correctly detected, respectively [37]. The F1-Score, presented in Equation 4 and calculated as a harmonic mean of precision and recall, integrates these two metrics to produce a single measure that balances their contributions, particularly valuable in situations with uneven class distributions [38].

$$Accuarcy = \frac{TP+TN}{TP+FN+FP+TN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1score = \frac{2*Recall*Precision}{(Recall+Precision)} \tag{4}$$

4. Results

A total of 12 different models were created using 4 different classifier algorithms and 3 different feature extractors. The results demonstrating the highest success across areas of interest have been presented in Table 5. The filtering column in the table shows the minimum and maximum word ratios excluded by each feature extractor. The first criterion mentioned indicates the number of words to be filtered based on the minimum occurrence, while the second criterion shows the ratio of words to be filtered based on their maximum occurrence. For instance, in the area of interest 'Terrorism', if a word appears 7 times or less, or more than 50% in the text content, it is removed during the model training process.

Table 5. Classification results.

Area of Interest	Classifier	Feature Extractor	acc	prec	rec	F1	Filtering (Min-Max)
Erdogan	RF	WTF	86.72	86.75	86.85	86.37	6 – 0.7
Refugees and Minorities	SVC	TF-IDF	89.74	89.85	90.55	89.52	5 – 0.6
International Relations	RF	TF-IDF	85.43	85.16	85.49	85.14	12 – 0.5
Sports and Tourism	KNN	TF-IDF	92.65	93.34	92.71	92.53	9 – 0.6
Elections	RF	TF-IDF	87.84	88.55	87.47	87.34	11 – 0.6
Natural Disaster and Epidemic	RF	TF-IDF	95.80	96.24	95.47	95.64	7 – 0.4
Terror	RF	TF-IDF	87.67	87.36	88.07	87.52	7 – 0.5
Economy	RF	TF-IDF	89.05	89.10	88.98	88.56	8 – 0.5

The model created for each area of interest was tested with numerous hyperparameters, and the optimal results are presented in Table 5. For the Random Forest algorithm, the number of trees was varied between 300 and 1500 in increments of 50. Both Gini and entropy were employed separately as criteria to measure the quality of the splits. For the SVC model, the kernels tested included linear, poly, RBF, and sigmoid. The degree of the polynomial kernel function was assessed from 3 to 30, with increments of 3. For the KNN model, the number of neighbors was tested from 4 to 20 in increments of 2. 'Uniform' and 'Distance' were evaluated separately as weight functions, and both Minkowski and Euclidean metrics were independently tested for distance computation. In this context, the necessary feature extractors, classifiers, and hyperparameters have been specified for each area of interest. Newly arriving news content will be processed through these individually determined models for each interest area, and will be labeled according to the class of the model that achieves the highest accuracy result.

The WTF feature extractor was used exclusively for the "Erdoğan" area of interest. For all other areas, the best results were achieved with the TF-IDF. For the "Erdoğan" area, the best results were obtained with a Random Forest (RF) algorithm running 1200 trees and using the "entropy" measure. For the areas of "Natural Disasters and Epidemics," "Terror," and "Elections," the best results were respectively achieved with RF algorithms using 1000, 900, and 1100 trees, all with the "entropy" measure. Similarly, for the "Economy" and "International Relations" areas, the best results were obtained with an RF algorithm running 1100 trees and using the "gini" measure. The "Refugees and Minorities" area achieved the best results with a 30-degree, linear Support Vector Classifier (SVC). For the "Sports and Tourism" area, a classifier using 18 neighbors, 'distance' weight calculation, and 'euclidean' measure was used.

Table 6. Correlation of news counts by areas of ınterest.

Area of Interest	**Erdogan**	**Refugees and Minorities**	**International Relations**	**Sports and Tourism**	**Elections**	**Natural Disaster and Epidemic**	**Terror**
Erdogan	1.00	0.62	0.78	0.42	0.52	0.17	0.77
Refugees and Minorities	0.62	1.00	0.77	0.17	0.21	0.01	0.83
International Relations	0.78	0.77	1.00	0.36	0.38	0.21	0.81
Sports and Tourism	0.42	0.17	0.36	1.00	0.33	0.45	0.15
Elections	0.52	0.21	0.38	0.33	1.00	0.72	0.22
Natural Disaster and Epidemic	0.17	0.01	0.21	0.45	0.72	1.00	0.08
Terror	0.77	0.83	0.81	0.15	0.22	0.08	1.00
Economy	0.46	0.27	0.64	0.61	0.57	0.62	0.22

In Figure 2, the relationship between the counts of news content, their publication dates, and their associated areas of interest is presented. This graph illustrates whether an increase in the number of news articles in one area of interest affects the news in other areas. Similarly, Table 6 demonstrates the correlation between different areas of interest based on the number of news articles published within the same time intervals.

Figure 3 visually represents the relationships and degrees of closeness between various areas of interest as derived from the analysis of news content. By mapping these relationships, the figure aids in identifying which areas frequently overlap in news reporting and which are distinctly segmented.

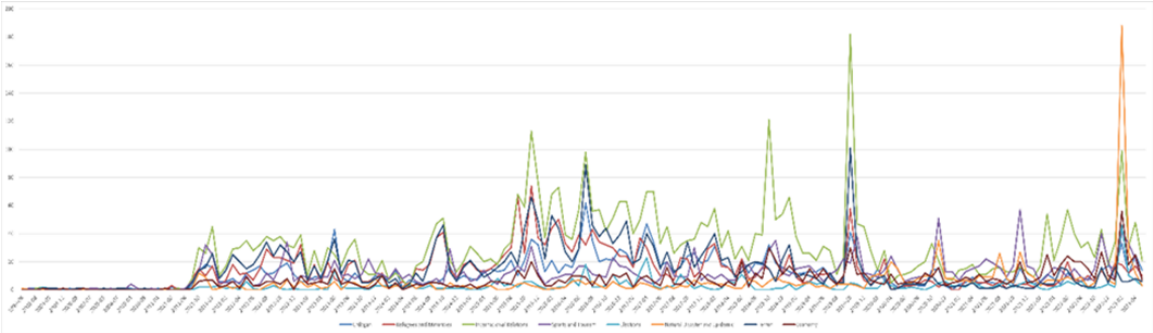

Figure 2. Relationship between news publication dates and areas of ınterest.

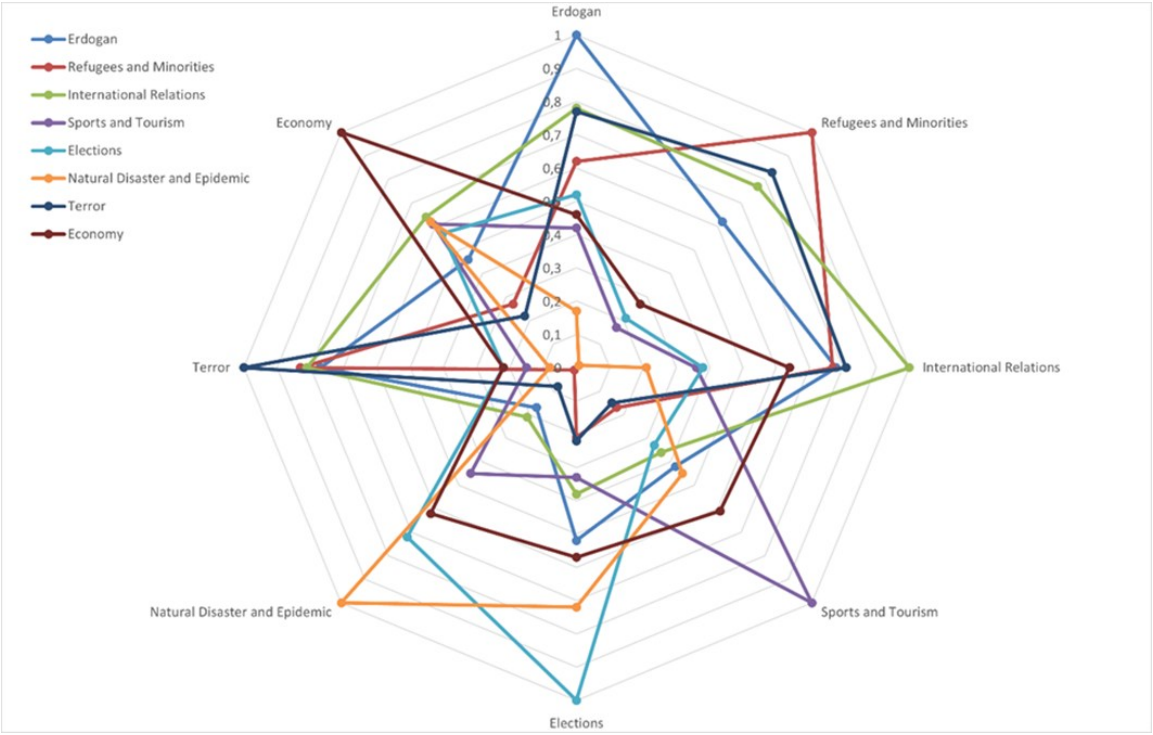

Figure 3. Proximity of areas of ınterest.

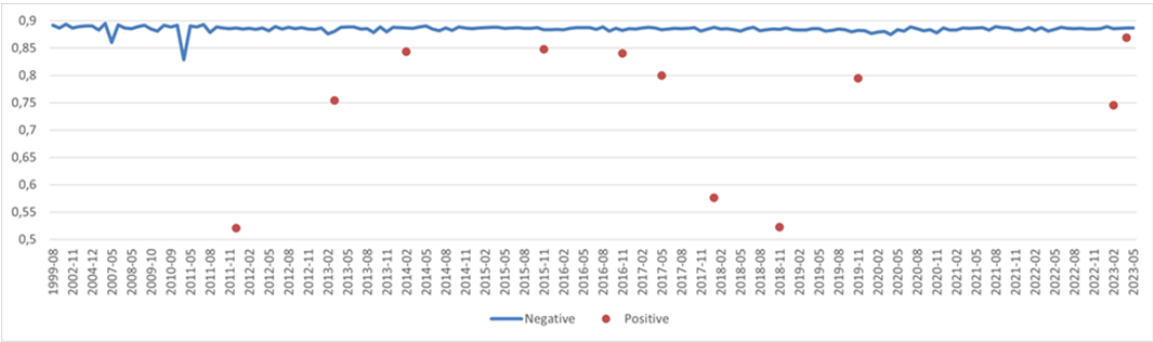

Figure 4. Analysis of BERT Sentiment Scores Over Time for News Content.

Table 7 displays the number of news articles in each area of interest along with the results obtained using the BERT algorithm. The table shows the counts of news content analyzed as negative or positive, and their average scores. The prevalence of such a high number of negative news articles in the BERT results appears as an adverse outcome. Consequently, irrespective of the area of interest, a random sample of 50 negative articles was manually read, and it was observed that even if they started positively or neutrally, they concluded with negative expressions.

Table 7. BERT Results by areas of ınterest.

Area of Interest	Number of News	Average of Negative BERT	Average of Positive BERT	Negative Number of News	Positive Number of News
Erdogan	1987	0.8855	0.7828	1984	3
Refugees and Minorities	2211	0.8868	0.7540	2210	1
International Relations	4811	0.8855	0.7041	4806	5
Sports and Tourism	1950	0.8822	0.7152	1945	5
Elections	482	0.8869	0.7997	481	1
Natural Disaster and Epidemic	915	0.8849	0.5228	914	1
Terror	2459	0.8870	0.6952	2457	2
Economy	1202	0.8836	0.7929	1200	2

Figure 4 provides a detailed visual analysis of the correlation between the sentiment scores derived from BERT and the publication dates of news content. By plotting BERT scores against the time of publication dates, the figure illustrates how perceptions of Turkey have shifted over time in the foreign press. According to the BERT algorithm, the number of news articles with positive sentiment is quite low, hence they are represented only as points on the graph. It is important to note that the BERT model was used in its pre-trained form without fine-tuning. Therefore, F1-score, which is typically calculated for trained models, is not applicable in this case. The BERT sentiment results presented here reflect direct classifications from the pre-trained model rather than performance metrics of a trained system.

5. Discussion

During the initial collection of news content, all articles containing the keywords "turkey," "türkiye," "Türkiye," "Turkey," "TÜRKİYE," and "TURKEY" were included in the dataset. However, in some of these news contents, the keyword appeared only once and the content might not be related to Turkey. Similarly, in some cases, the searched keyword could have different meanings, such as being the name of a dish. These scenarios have negatively impacted the classification success of news content according to areas of interest.

In the classification of news content according to areas of interest, separate filtering criteria, classification algorithms, and feature extractors were used due to the considerably different textual concepts present in each area. The reason for this variation in filtering rates stems from the distinctly different word structures contained in the news text according to the area of interest. Upon reviewing Table 5, it appears that the

"International Relations" area has the lowest minimum ratio. This is designed to counteract the dominance of country names in news content associated with this area of interest. Similarly, the highest maximum ratio was applied in the "Natural Disasters and Epidemics" area. The rationale behind this choice is to prevent frequently occurring words such as earthquake, COVID, pandemic, aftershock, etc., from being decisive in classification processes.

The highest classification success in the news content was achieved in the "Natural Disasters and Epidemics" area, followed by the "Sports and Tourism" area. The reason for this is the frequent use of unique words specific to these fields. The lowest success was observed in the "International Relations" area and subsequently in the "Erdoğan" area. One reason for this is that news content in these areas often consists of very general expressions. All news content was classified according to predefined areas of interest with an average accuracy rate of 89.36%.

The graph presented in Figure 2 reveals intriguing details about the relationship between the timing of news publication and the contents. The most frequently labeled news content pertains to the "International Relations" area of interest. This prevalence is expected since the web pages from which the news texts are collected are international content providers. Subsequently, the majority of news concerning Turkey has been related to terrorism. The peaks in news volume about terrorism have been shaped according to the Turkish news agenda. For example, in 2019, the most frequent period for terrorism news was in October/November, coinciding with Turkey's initiation of the "Operation Peace Spring." During these months, news worldwide concerning Turkey and terrorism reached its highest level. Another peak in terrorism-related news occurred in July/August 2016, which corresponds to the coup attempt in Turkey and its aftermath. Similarly, the "Natural Disasters and Epidemics" area saw the most news coverage during the onset of the COVID pandemic and the dates of significant earthquakes. It can be concluded that the proposed model demonstrates the effectiveness of classifying news content according to areas of interest.

When examining the correlation relationships between the numbers of news articles in various areas of interest based on publication dates (Table 6 and Figure 3), it is evident that some areas exhibit a high degree of correlation. The highest correlation exists between the "Refugees and Minorities" and "Terror" areas of interest. The next highest correlation is between the "International Relations" and "Terror" areas. This indicates that the numbers and dates of news articles published in the "Terror" area are closely linked with those in the "Refugees and Minorities" and "International Relations" areas. This situation demonstrates that news related to terrorism is directly associated with news concerning "International Relations" and "Refugees and Minorities." In other words, an increase in the number of news articles in one of these areas directly affects the other. Additionally, the "Sports and Tourism" area has almost no correlation with any other area, which is likely due to the nature of the content in this area being significantly different from others. The lowest correlation is between the "Natural Disasters and Epidemics" area and the "Refugees and Minorities" area, attributed to the very low relational linkage between news in these areas.

Upon examining Table 7 and Figure 4, it is apparent that the sentiment analysis results conducted on news content using the BERT algorithm are predominantly negative. The number of news articles with positive sentiment is negligibly small, and the scores of

those few positive cases are quite low. This is true even for news in the "Sports and Tourism" areas of interest. In this context, it can be asserted that almost all news related to Turkey is perceived negatively. Upon reviewing the news content, it has been observed that even content normally expected to be positive eventually changes tone, incorporating negative sentences that alter the sentiment of the piece.

Table 8. Related works and presented study.

Study	Purpose	Method
Hassani and colleagues [3]	To explore the potential applications of text mining in big data analytics.	Suggestions for successful news content analysis using text mining.
Leonard and colleagues [11]	To classify news content based on headlines.	SVC model achieved 90% accuracy compared to LR, MNB, Decision Tree, and RF.
Agarwal and colleagues [12]	To classify news based on areas of interest.	RF, Decision Tree, KNN, Gaussian Naive Bayes; RF achieved 91.94% accuracy.
Sadjadi and colleagues [13]	To classify news using both labelled and unlabeled data.	Semi-supervised model with accuracy of 77.3% to 81.92%.
Miao and colleagues [14]	To classify Chinese news content.	TF-IDF + KNN, NB, SVM; highest F1-Score with SVM at 95.7%.
Salehin and associates [15]	To classify Bangla news based on publisher's categories.	Multiple ML models; LSTM achieved 87% accuracy.
Lin and colleagues [16]	To detect harmful news using sentiment analysis.	BERT-based model with F1-score of 66.3%.
Hayawi and colleagues [17]	To perform sentiment analysis on COVID-19 news.	XGBoost, LSTM, BERT; BERT achieved 98% accuracy.
Ahmet et al. [19]	To review approaches for online news classification based on publisher's categories.	Bayesian classifier with 93% accuracy.
Proposed Study	To classify news based on areas of interest and analyze Turkey's image in the foreign press through sentiment analysis	BERT for sentiment, RF, KNN, SVM, MNB for classification

To clarify the contributions of this study and distinguish it from previous works, Table 8 presents a comparative analysis of related studies in the field of text mining, news classification, and sentiment analysis. While prior research has primarily focused on categorizing news content based on predefined publisher-driven classifications, this study offers a novel approach by integrating sentiment analysis with topic-based classification specific to Turkey's image in the foreign press. The table highlights the objectives, methodologies, and key findings of existing works, emphasizing the unique aspects of our study and how it addresses the gaps in the literature.

In the literature, the primary use of the BERT algorithm on news content has been observed for the purpose of detecting fake news. A conclusion drawn from similar studies in this area is that sentiment analysis of news content using BERT is quite successful.

Many studies related to the classification of news content are also evident. However, it has generally been found that these studies classify news content according to categories provided by the publisher. In this study, however, news content has been grouped according to areas of interest. The categorization provided by the publisher and the areas of interest convey different meanings in the grouping of news content. While the category typically shows the subject of the news content, the area of interest represents the significance conveyed by the content rather than just the subject. For example, a news item labeled "International Relations" would cover stories about Turkey's relations with other countries. Thus, a sports news item would be included in this area if it is of an international level, but not if it is on a local level. Similarly, an economic content piece would have the "International Relations" tag if it is based on relations between Turkey and other countries or concerns the economy of other countries.

The predominance of negative sentiment in news content as identified by BERT suggests a potential bias in foreign media's portrayal of Turkey. This finding aligns with previous studies indicating that international news coverage often emphasizes conflict-oriented narratives. Addressing this bias in future research could involve expanding the dataset to include more neutral sources and exploring advanced sentiment analysis techniques.

6. Conclusion

In this study, a comprehensive dataset consisting of 8,385 news contents related to Turkey was collected from four different sources and shared online for other researchers' use [35]. The purpose of the study is to analyze the perception of Turkey in the foreign press and to examine the relationships between the publication dates of news contents across different areas of interest. In this context, the study focused on classifying news contents according to their areas of interest, extracting relationships based on their publication dates, and performing sentiment analysis using the BERT algorithm. During the digitization process of the news contents, three different feature extractors—TF-IDF, WIDF, and GSW—were used, and various classifiers such as RF, KNN, MNB, and SVM were employed in the model development process. Considering each area of interest independently, models that would yield the best results were created using different feature extractors, classifiers, and hyperparameters for each area. An average accuracy of 89.36% was achieved in the classification of news contents according to their areas of interest. The sentiment results from the BERT analysis of almost all news contents were found to be negative. Intriguing correlations and connections have been discovered between the publication dates of news contents belonging to different areas of interest.
In subsequent phases, the study can be expanded to include more datasets. By analyzing news content from different countries, the changes in the perception of Turkey in the foreign press over the years can be comparatively studied across various nations.

While the study provides valuable insights into the portrayal of Turkey in foreign media, it is limited by the scope of its dataset and the predominance of negative sentiment. Future research could focus on expanding the dataset to include more diverse sources and applying alternative sentiment analysis models to validate the findings. Additionally, exploring temporal changes in sentiment could further clarify trends in media portrayal.

References

[1] Austin, C., and Kusumoto, F., The application of Big Data in medicine: current implications and future directions, **Journal of Interventional Cardiac Electrophysiology**, **47**, 51-59, (2016)

[2] Rana, M.I., Khalid, S., and Akbar, M.U., News classification based on their headlines: A review, **17th IEEE International Multi Topic Conference 2014**, (2014)

[3] Hassani, H., Beneki, C., Unger, S., Mazinani, M.T., and Yeganegi, M.R., Text mining in big data analytics, **Big Data and Cognitive Computing**, **4**, 1, 1, (2020)

[4] Kaur, G., and Bajaj, K., News classification and its techniques: a review, **IOSR Journal of Computer Engineering**, **18**, 1, 22-26, (2016)

[5] Dadgar, S.M.H., Araghi, M.S., and Farahani, M.M., A novel text mining approach based on TF-IDF and Support Vector Machine for news classification, **2016 IEEE International Conference on Engineering and Technology (ICETECH)**, (2016)

[6] Ghomi, H., and Hussein, M., An integrated text mining, literature review, and meta-analysis approach to investigate pedestrian violation behaviours, **Accident Analysis & Prevention**, **173**, 106712, (2022)

[7] Gomes, H., de Castro Neto, M., and Henriques, R., Text Mining: Sentiment analysis on news classification, **2013 8th Iberian Conference on Information Systems and Technologies (CISTI)**, (2013)

[8] Zhang, X., and Li, W., From social media with news: Journalists' social media use for sourcing and verification, **Journalism Practice**, **14**, 10, 1193-1210, (2020)

[9] Cetina Presuel, R., and Sierra, J.M.M., Algorithms and the news: social media platforms as news publishers and distributors, **Revista De Comunicación**, **18**, 2, 261-285, (2019)

[10] Carreira, R., Crato, J.M., Gonçalves, D., and Jorge, J.A., Evaluating adaptive user profiles for news classification, **Proceedings of the 9th international conference on Intelligent user interfaces**, (2004)

[11] Leonard, G., Sisnadi, F., Wardhana, N.V., Al-Ghofari, M.A.A., and Girsang, A.S., News Classification Based On News Headline Using SVC Classifier, **2022 16th International Conference on Telecommunication Systems, Services, and Applications (TSSA)**, (2022)

[12] Agarwal, J., Christa, S., Pai, A., Kumar, M.A., and Prasad, G., Machine learning application for news text classification, **2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence)**, (2023)

[13] Sadjadi, S., Mashayekhi, H., and Hassanpour, H., A two-level semi-supervised clustering technique for news articles, **International Journal of Engineering**, **34**, 12, 2648-2657, (2021)

[14] Miao, F., Zhang, P., Jin, L., and Wu, H., Chinese news text classification based on machine learning algorithm, **2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)**, (2018)

[15] Salehin, K., Alam, M.K., Nabi, M.A., Ahmed, F., and Ashraf, F.B., A comparative study of different text classification approaches for bangla news classification, **2021 24th International Conference on Computer and Information Technology (ICCIT)**, (2021)

[16] Lin, S.-Y., Kung, Y.-C., and Leu, F.-Y., Predictive intelligence in harmful news identification by BERT-based ensemble learning model with text sentiment analysis, **Information Processing & Management**, **59**, 2, 102872, (2022)

[17] Hayawi, K., Shahriar, S., Serhani, M.A., Taleb, I., and Mathew, S.S., ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection, **Public health**, **203**, 23-30, (2022)

[18] To, Q.G., To, K.G., Huynh, V.-A.N., Nguyen, N.T., Ngo, D.T., Alley, S.J., Tran, A.N., Tran, A.N., Pham, N.T., and Bui, T.X., Applying machine learning to identify anti-vaccination tweets during the COVID-19 pandemic, **International journal of environmental research and public health**, **18**, 8, 4069, (2021)

[19] Ahmed, J., and Ahmed, M., Online news classification using machine learning techniques, **IIUM Engineering Journal**, **22**, 2, 210-225, (2021)

[20] MAHAJAN, S., and Ingle, D., News classification using machine learning, **Int. J. Recent Innov. Trends Comput. Commun**, **9**, 5, 23-27, (2021)

[21] Sunagar, P., Kanavalli, A., Nayak, S.S., Mahan, S.R., Prasad, S., and Prasad, S., News Topic Classification Using Machine Learning Techniques, **International Conference on Communication, Computing and Electronics Systems: Proceedings of ICCCES 2020**, (2021)

[22] Fanny, F., Muliono, Y., and Tanzil, F., A comparison of text classification methods k-NN, Naïve Bayes, and support vector machine for news classification, **Jurnal Informatika: Jurnal Pengembangan IT**, **3**, 2, 157-160, (2018)

[23] Nwet, K.T., and Darren, S., Machine learning algorithms for Myanmar news classification, **International Journal on Natural Language Computing (IJNLC)**, **8**, 4, (2019)

[24] Keya, A.J., Wadud, M.A.H., Mridha, M., Alatiyyah, M., and Hamid, M.A., AugFake-BERT: handling imbalance through augmentation of fake news using BERT to enhance the performance of fake news classification, **Applied Sciences**, **12**, 17, 8398, (2022)

[25] Shishah, W., Fake news detection using BERT model with joint learning, **Arabian Journal for Science and Engineering**, **46**, 9, 9115-9127, (2021)

[26] Mehta, D., Dwivedi, A., Patra, A., and Anand Kumar, M., A transformer-based architecture for fake news classification, **Social network analysis and mining**, **11**, 1-12, (2021)

[27] Sparck Jones, K., A statistical interpretation of term specificity and its application in retrieval, **Journal of documentation**, **28**, 1, 11-21, (1972)

[28] Takenobu, T., Text categorization based on weighted inverse document frequency, **Information Processing Society of Japan, SIGNL**, **94**, 100, 33-40, (1994)

[29] Sabbah, T., Selamat, A., Selamat, M.H., Al-Anzi, F.S., Viedma, E.H., Krejcar, O., and Fujita, H., Modified frequency-based term weighting schemes for text classification, **Applied Soft Computing**, **58**, 193-206, (2017)

[30] Breiman, L., Random forests, **Machine learning**, **45**, 5-32, (2001)

[31] Boser, B.E., Guyon, I.M., and Vapnik, V.N., A training algorithm for optimal margin classifiers, **Proceedings of the fifth annual workshop on Computational learning theory**, (1992)

[32] Fix, E., and Hodges, J.L., Discriminatory analysis, nonparametric discrimination, (1951)

[33] McCallum, A., and Nigam, K., A comparison of event models for naive bayes text classification, **AAAI-98 workshop on learning for text categorization**, (1998)

[34] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K., Bert: Pre-training of deep bidirectional transformers for language understanding, **arXiv preprint arXiv:1810.04805**, (2018)

[35] ISIK, M., and AYDEMİR, E., News about Turkey from BBC, CNN, TG, and FOX. (2024)

[36] Fawcett, T., An introduction to ROC analysis, **Pattern recognition letters**, **27**, 8, 861-874, (2006)

[37] Powers, D.M., Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, **arXiv preprint arXiv:2010.16061**, (2020)

[38] Takahashi, K., Yamamoto, K., Kuchiba, A., and Koyama, T., Confidence interval for micro-averaged F 1 and macro-averaged F 1 scores, **Applied Intelligence**, **52**, 5, 4961-4972, (2022)