Journal of Universal Computer Science, vol. 32, no. 4 (2026), 615-643
submitted: 29/5/2025, accepted: 23/9/2025, appeared: 28/4/2026

Duygu-Turk: A Context-Aware Sentiment Analysis Framework for Turkish, Based on Plutchik's Emotion Model

Rabia Tintin
(Canakkale Onsekiz Mart University, Canakkale, Turkey
https://orcid.org/0009-0002-2257-2360, rabiatintin@comu.edu.tr)

Sait Can Yucebas
(Canakkale Onsekiz Mart University, Canakkale, Turkey
https://orcid.org/0000-0002-1030-3545, can@comu.edu.tr)

Abstract: This study presents Duygu-Turk, a novel deep learning-based sentiment analysis framework specifically designed for the Turkish language which is characterized by its agglutinative and morphologically rich structure. Unlike conventional sentiment analysis models that rely on coarse polarity classification (positive, negative, neutral) and insufficient integration of Turkish-specific linguistic features, Duygu-Turk adopts a fine-grained classification approach based on Plutchik's Wheel of Emotions. The model identifies eight primary emotions, eight secondary emotions, and varying degrees of emotional intensity. Additionally, a non-monotonic logic mechanism is integrated to detect conditional sentiments, allowing for more context-sensitive classification. To enhance linguistic coverage, the model leverages morpho-semantic features, idiomatic expressions, suffixes, and contrastive conjunctions unique to Turkish. A new sentiment corpus consisting of 136,000 annotated Turkish sentences was constructed to train and validate the model. Experimental evaluations demonstrate that Duygu-Turk significantly outperforms transformer-based models such as BERT, DistilBERT, and ELECTRA, achieving F1 scores of 0.99 for polarity classification and 0.90 for multi-class emotion classification. These results highlight the model's potential as a robust and linguistically grounded solution for sentiment analysis in Turkish and other low-resource languages.

Keywords: Sentiment analysis, Natural language processing, LSTM, Turkish language
Categories: I.2.1, I.2.7, I.2.6, I.5.1
DOI: 10.3897/jucs.160588

1 Introduction

The rapid advancements in natural language processing (NLP) and the increasing volume of user-generated content on social media platforms have underscored the growing importance of sentiment analysis, particularly for morphologically rich and low-resource languages such as Turkish. Sentiment analysis not only facilitates the extraction of subjective information but also supports informed decision-making across various domains, including marketing, public policy, and healthcare [Wankhade, 22].

In addition to these applications, sentiment analysis plays a critical role in combating misinformation and curbing the dissemination of misleading content—one of today's most pressing challenges [Alonso, 21]. Specifically, sentiment analysis techniques can be leveraged to assess the emotional tone of online content, thereby enabling the evaluation of its credibility and helping to differentiate between reliable and deceptive information. Such analysis can also enhance content moderation systems, improve digital literacy, and assist in determining the appropriateness of content for specific audiences. For example, analyzing the sentiment conveyed in book reviews or summaries can aid in evaluating whether the material is suitable for children.

Despite the increasing relevance of sentiment analysis, existing research has largely focused on high-resource languages such as English [Aydın, 21], while Turkish remains comparatively underexplored [Balli, 22]. The studies conducted on Turkish sentiment analysis to date have primarily adopted two methodological paradigms: lexicon-based approaches and transformer-based models. Early lexicon-based methods [Vural, 13], [Demirtas, 13] provided foundational resources for the field, but they often suffer from limited lexical coverage and inadequate handling of the semantic variability introduced by Turkish's agglutinative structure. On the other hand, transformer models such as BERT have recently gained prominence due to their ability to model contextual dependencies. While these models have demonstrated strong performance in high-resource settings, their use in Turkish sentiment analysis remains constrained by several limitations, including reduced interpretability, context loss in longer or complex sentences, and high dependence on large-scale labeled datasets and computational resources [Alawi, 24].

Moreover, an in-depth review of Turkish sentiment analysis literature reveals that the majority of studies—regardless of the method employed—have relied on simplified sentiment polarity classification, typically using only two (positive/negative) or three (positive/negative/neutral) categories. An important dimension of sentiment, namely emotional intensity, has been largely overlooked [Balli, 22], [Uyaroğlu Akdeniz, 21], [Aslan, 23], [Polat, 22], [Shehu, 21]. In addition, essential linguistic features unique to Turkish—such as idioms, proverbs, compound verbs, formulaic expressions, and morpho-semantic suffixes—have received little attention, thereby limiting the accuracy and cultural relevance of current sentiment models.

Although a small number of studies have proposed more than three emotion categories [Güven, 21], [Alkan, 23], these typically involve a maximum of five classes, often selected without a clear theoretical grounding. To the best of our knowledge, there has been no attempt to identify compound or secondary emotions that arise from the co-occurrence of multiple primary emotions—a critical gap in modeling the full spectrum of human affect.

To address these limitations, this study introduces Duygu-Turk, a deep learning-based sentiment analysis framework tailored specifically for Turkish. Drawing on Plutchik's wheel of emotions [Plutchik, 23], the model is designed to classify sentiment across 16 categories, each associated with varying levels of intensity. Unlike conventional three-way classification models, Duygu-Turk offers a nuanced, fine-grained sentiment categorization that enhances interpretability and expressiveness.

Furthermore, Duygu-Turk incorporates non-monotonic logic to detect conditionally positive and conditionally negative sentiment states, allowing for more context-aware analysis. The model also integrates key linguistic characteristics of Turkish, including morpho-semantic suffixes, compound verbs, idiomatic expressions, and context-dependent emotional cues. This linguistically informed and context-sensitive approach enables Duygu-Turk to overcome the limitations of existing models, offering a more flexible, accurate, and resource-efficient solution for sentiment analysis in Turkish. In doing so, the proposed methodology not only strengthens the representational capacity of sentiment analysis systems but also promotes broader applicability in real-world, low-resource NLP contexts.

2 Related Studies

Uyaroğlu Akdeniz and Cebeci [Uyaroğlu Akdeniz, 21] analyzed Twitter comments regarding municipal services in Sakarya. In the study, user comments were assigned to five categories based on different municipal service areas. Naïve Bayes (NB), Decision Tree (DT), Gradient Boosting Tree (GBT), and Random Forest (RF) classifiers were used for this multi-class classification task. Among these methods, NB achieved the highest accuracy of 92%. For sentiment detection within each category, the Turkish BERT model was employed. However, the study did not report any performance metrics for the Turkish BERT model. Rather than proposing a new method, the study focused on applying the widely adopted BERT model to Turkish text. Sentiment classification was limited to two categories: positive and negative. Neither sentiment types nor sentiment intensity levels were analyzed. Furthermore, the study considered only word stems, excluding derivational and inflectional suffixes. Given that Turkish is an agglutinative language, this limitation may have led to the omission of derived words with distinct meanings, potentially affecting the accuracy of the sentiment analysis.

In the study by Bozuyla [Bozuyla, 24], the performance of traditional machine learning (ML) methods, deep learning approaches, and transformer-based models (BERT, RoBERTa, DistilBERT, BERTurk) was compared over Turkish drug review dataset. During the preprocessing stage, traditional ML algorithms employed Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF) and Zemberek. For deep learning models, Word2Vec was utilized, whereas transformer models employed tokenization, padding, and mask creation. The BERTurk model achieved the highest performance with an F1 score of 95.1%. The study considered only three basic sentiment classes: positive, negative, and neutral. It did not examine different types of sentiments or their intensities. This highlights the need to expand the scope of sentiment analysis.

In their study, Tohma et al. [Tohma, 23], proposed a BERT-based model to enhance the performance of sentiment analysis in Turkish question answering systems. Preprocessed question-answer texts, along with a pre-trained Turkish BERT model, were used. Additionally, TF-IDF and Word2Vec were employed to convert texts into vector form. The dataset also included pre-labeled polarity vectors containing positive and negative word scores. The proposed method achieved an accuracy of 91%. This approach differentiates itself by creating a polarity vector that incorporates tone scores into word vectors. This approach improved the performance of the BERT model. However, it requires that all texts be pre-labeled with polarity scores, which is very time-consuming and inefficient.

Tepecik and Demir [Tepecik, 24] performed sentiment analysis on transcribed Turkish speech recordings. BERT and ELECTRA models were used to detect three emotion states. The data were converted into vectors using CountVectorizer and TF-IDFVectorizer methods. For the accuracy analysis of emotion detection, Naive Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) algorithms were employed. The results showed that the combination of the BERT model with NB and LR algorithms achieved the highest accuracy of 70%. Similarly, the combination of the ELECTRA model with RF and LR algorithms both achieved the highest accuracy of 72%. However, the accuracy rates in this study are not at the desired level, and, as in many studies, only three emotion classes were considered.

Guven's study [Güven, 21] distinguishes itself from other research for Turkish sentiment analysis by increasing the number of sentiment classes. The study compares the performance of BERT-based models (BERT, Bert-T, Bert-M, Dbert-T) with ML methods in classifying five sentiments: angry, happy, sad, surprised, and fearful. For ML, Random Forest (RF), Naïve Bayes (NB), and Logistic Regression (LR) were employed. Among these, LR achieved the highest accuracy at 98.4%, while Bert-T emerged as the most effective overall method with 98.75% accuracy. Despite being one of the few studies to explore an expanded set of sentiment classes, the criteria for selecting these sentiments are not clearly explained. Furthermore, sentiment intensity is not analyzed. From a methodological perspective, deep learning models that have demonstrated high success in sentiment analysis are not sufficiently explored.

Polat and Ağca [Polat, 22] analyzed sentiment trends in Turkish and English TripAdvisor reviews, classifying them into positive and negative categories. The study compares lexicon-based methods with ML approaches. One limitation of the study is that it only considers two sentiment categories, without accounting for sentiment types or intensities. While lexicon-based methods showed strong performance on the English dataset, ML methods outperformed them on the Turkish dataset. This finding underscores the limitations of existing sentiment lexicons for the Turkish language. It also highlights the crucial role of ML methods in Turkish sentiment analysis.

Balli et al. [Balli, 22] used two different datasets in their study. The first dataset consisted of general tweets, while the second was the manually created SentimentSet dataset. SentimentSet was constructed by manually labeling Turkish tweets filtered by keywords such as "pandemic" and "corona" as positive, negative, or neutral. During the preprocessing step, ML methods were tested using the Zemberek and Snowball algorithms. In the second phase, the LSTM model was tested without preprocessing. Results from all methods were then compared. In the SentimentSet dataset, the model using Zemberek preprocessing and SVM achieved an accuracy of 87.47%, while LSTM without preprocessing showed an accuracy of 86.30%. This finding highlights the positive impact of data preprocessing in sentiment analysis tasks. However, despite the availability of positive, negative, and neutral labeled tweets in the dataset, the classification was conducted only between positive and negative classes. Additionally, the LSTM model showed lower performance compared to traditional ML methods.

Shehu et al. [Shehu, 21] compared traditional ML algorithms with deep learning algorithms for Turkish sentiment analysis. To expand the training dataset, they employed three data augmentation techniques: Shift, Shuffle, and Hybrid. In the preprocessing step, word stems were extracted, and sentiment analysis was conducted using RNN, CNN, and HAN models. The performance of these deep learning models was compared with traditional machine learning models, including RSVM, RF, ME, SVM, and DT. While traditional machine learning methods demonstrated efficiency in terms of computational cost, deep learning models outperformed them in classification accuracy. The most significant contribution of the study is its investigation of the impact of stemming on learning algorithm performance. The findings indicate that stemming positively influences the performance of deep learning models, leading to an 8.8% increase in accuracy. This improvement highlights the substantial role of stemming in enhancing deep learning models' effectiveness. However, as in most existing studies, only two sentiment classes were considered, disregarding variations in sentiment types and intensity.

Savci and Das [Savci, 23] conducted a comparative sentiment analysis on texts in Turkish, English, and Arabic by constructing e-commerce datasets for these languages. Their study compared a pre-trained language model with a deep learning model. For pre-trained models, they used NB, DVM, and RO, while deep learning methods included LSTM, recurrent neural networks (RNN), and convolutional neural networks (CNN). The models were evaluated based on their ability to classify sentiments into three categories: positive, negative, and neutral. Among the pre-trained models, DVM achieved the highest accuracy for all three languages. Among deep learning models, RNN performed best, regardless of language. When analyzing performance across languages, both pre-trained and deep learning models achieved higher accuracy in English. These findings highlight the need for further research to improve sentiment analysis in the Turkish language.

Uca et al. [Uca, 22] conducted sentiment classification of healthcare data. The researchers employed the LSTM model for sentiment analysis, highlighting its effectiveness in processing long texts with a reported performance score of 94%. As in similar studies, sentiments were categorized into three classes: positive, negative, and neutral. Additionally, since the study focuses solely on healthcare data, it presents a domain-specific approach rather than a general sentiment analysis framework.

Aslan [Aslan, 23] proposed a feature-based sentiment analysis approach. This approach categorizes various opinions in user reviews to generate more targeted recommendations. In the study, the most frequently used words in English user reviews were identified using the TF-IDF. For each detected word, semantically related terms were determined using Word2Vec. Sentiment polarity was then analyzed using the VADER tool, which relies on a predefined lexicon to assign positive, negative, and neutral sentiment scores to words. The study found that sentiment distributions varied across different features. However, no performance evaluation of the proposed method was provided. Lexicon-based approaches have two major drawbacks. First, there is a high likelihood of significant performance degradation when encountering words that are not present in the lexicon. Since the model relies on a predefined word list, it struggles to analyze terms outside its vocabulary. Second, these approaches evaluate words without considering their contextual meaning, which can lead to misinterpretations and further performance loss [Sağlam, 19].

Methodological approaches in recent studies indicate the prevalence of three primary techniques: lexicon-based methods, transformer models, and ML. Lexicon-based methods are limited by their dependence on predefined lexicons. Although transformer models, such as BERT, achieve high performance, they suffer from context loss in long texts [Kaya, 24]. Additionally, for Turkish, the complexity of morphological structures and the lack of comprehensive datasets pose challenges for transformer-based models. Studies on Turkish sentiment analysis have demonstrated the effectiveness of ML methods, with LSTM standing out due to its scalability, low computational cost, and ability to retain contextual information in long sentences [Ozturk, 22].

Independent of the methodological approach, most sentiment analysis studies have been conducted in English, while research on the Turkish language remains limited [Balli, 22]. These studies primarily classify sentiments into two (positive, negative) or three categories (positive, negative, neutral). Sentiment intensity, which indicates the strength of sentiment, has also been largely overlooked [Balli, 22], [Uyaroğlu Akdeniz, 21], [Aslan, 23], [Polat, 22], [Shehu, 21]. Furthermore, elements specific to Turkish, such as proverbs, idioms, compound verbs, fixed expressions, and morpho-semantic affixes, have not been adequately considered in sentiment analysis. In addition, only a few studies argue that sentiment should be represented with more than three classes. Among them, the highest number of sentiment categories identified is five [Güven, 21], [Alkan, 23]. However, these studies arbitrarily define sentiment categories without a strong theoretical basis. Additionally, to our knowledge, no existing study has attempted to identify secondary (intermediate) sentiments that emerge from the combination of multiple primary sentiments.

To address these limitations, Duygu-Turk, a comprehensive sentiment analysis model incorporating Turkish-specific linguistic structures and sentiment intensity, has been developed. Duygu-Turk is based on Plutchik's wheel of emotions [Plutchik, 23], a widely used model in psychology for identifying primary and secondary emotions. Unlike existing studies, Duygu-Turk not only detects positive, negative, and neutral sentiments but also identifies conditionally positive and conditionally negative states, the eight primary emotions from Plutchik's model, their intensities, and secondary emotions that emerge from their intersections

3 Conceptual Framework

The majority of current sentiment analysis studies [Balli, 22], [Uyaroğlu Akdeniz, 21], [Aslan, 23], [Polat, 22], [Shehu, 21], [Bozuyla, 24], [Tohma, 23], [Tepecik, 24], [Savci, 23], [Uca, 22] evaluate sentiments by classifying them into two or three categories as positive, negative, and neutral. In addition, studies that utilize more than three sentiment categories are relatively limited [Güven, 21]. Furthermore, it is noteworthy that these studies often lack clear criteria for the selection of emotions.

This study is the first in the literature to apply Plutchik's Wheel of Emotions to the Turkish language. Duygu-Turk utilizes this model to expand the set of primary emotions, determine emotion intensity, and label secondary emotions formed by emotion combinations.

The emotion theory developed by Plutchik [Plutchik, 23] argues that humans experience a set of fundamental emotions throughout their lives. To represent these primary emotions, variations in emotion intensity, and secondary emotions formed by their intersections, he developed the Wheel of Emotions model. This model is presented in Figure 1.

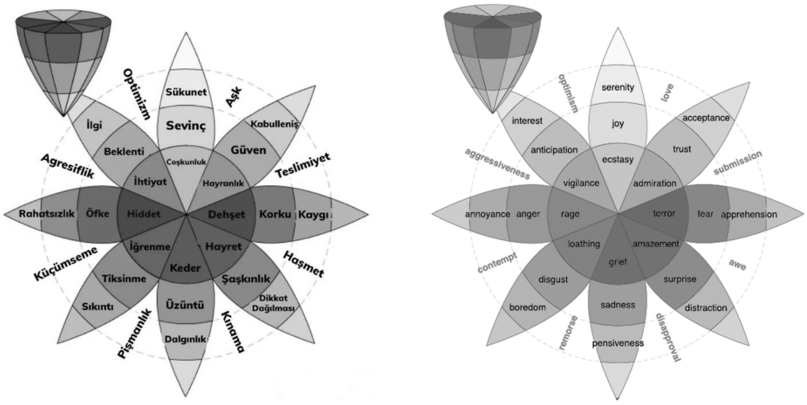

Figure 1: Plutchik's Wheel of Emotions [Plutchik, 23]

The eight primary emotions and their descriptions are as follows:

- Joy: Represents positive emotions such as success, happiness, love, and delight.
- Trust: Includes feelings of security, safety, peace, and comfort.
- Fear: Encompasses negative emotions like fear, anxiety, worry, and terror.
- Surprise: Covers neutral emotions such as astonishment, amazement, curiosity, and shock.
- Sadness: Represents negative emotions like loss, pain, loneliness, and sorrow.
- Disgust: Includes emotions such as repulsion, disgust, hatred, and aversion.
- Anger: Encompasses negative emotions like rage, irritation, wrath, and resentment.
- Anticipation: Represents neutral emotions such as expectation, hope, desire, and imagination.

These primary emotions interact to form secondary emotions that are listed below [Plutchik, 23]:

- Joy + Trust = Love
- Trust + Fear = Submission
- Fear + Surprise = Awe
- Surprise + Sadness = Disapproval
- Sadness + Disgust = Remorse
- Disgust + Anger = Contempt
- Anger + Anticipation = Aggressiveness
- Anticipation + Joy = Optimism

In this study, Duygu-Turk adopts Plutchik's Wheel of Emotions to identify the eight primary emotions, their combinations that form secondary emotions, and the varying levels of intensity associated with each.

4 Material and Methods

This study develops a sentiment analysis model based on Plutchik's Wheel of Emotions for Turkish written texts. The model is capable of identifying the eight primary emotions (Joy, Trust, Fear, Surprise, Sadness, Disgust, Anger, Anticipation), eight secondary emotions formed by their intersections (Optimism, Love, Submission, Awe, Disappointment, Remorse, Contempt, Aggressiveness), and the intensity of these emotions. The flow diagram of the model is presented in Figure 2.

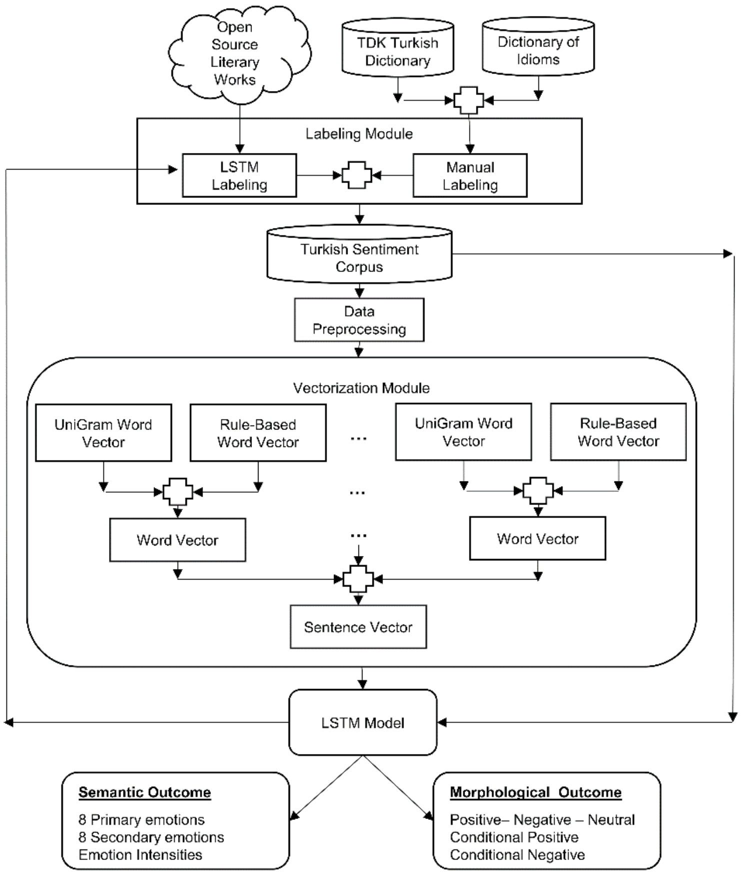

Figure 2: Flow diagram of Duygu-Turk model

A large-scale Turkish sentiment dataset was required for training the model. However, due to the insufficient scale of existing works in the literature, the first step of the study involved the development of a Turkish sentiment corpus. For this purpose, sentiment-labeled examples were manually and automatically tagged from various sources. After the preprocessing step, the data in the corpus is given to the Vectorization Module. In this module, for each example sentence, UniGram and Rule-Based word vectors are obtained. Then, these vectors are combined to form a sentence vector. Sentence vectors are then passed to the LSTM for both morphological and sentiment analysis. The details of the flow presented in Figure 2 are outlined in the following subsections.

4.1 Material

An examination of Turkish sentiment analysis studies reveals that the available lexicons and datasets are limited in number and have some content-related gaps. In terms of quantity, the most comprehensive dataset is the one created by Kumas [Kumaş, 21] with 32,000 tweets. However, the examples in this dataset include only two sentiment labels (positive and negative). Other studies used smaller datasets, such as Tocoglu and Alpkocak [Tocoglu, 18] with 27,350 entries, Güven et al. [Güven, 19] with 4,000, and Çetin and Eryiğit [Çetin, 18] with 1,415 entries. From a contextual perspective, most datasets have been created using a monotonic logic approach, which fails to sufficiently capture a range of emotions. The major deficiencies in these datasets are the lack of sufficient examples for emotions, the inability to distinguish homonymous words based on their contextual meaning, the exclusion of idiomatic expressions and proverbs containing sentiment, and the neglect of morpho-semantic structures.

To address these shortcomings and create a comprehensive corpus that can be used both as a lexicon and as a comprehensive dataset for Turkish sentiment analysis, a new corpus has been developed. While the development of this new corpus is not the primary objective of the study, it significantly contributes to the literature by enriching the existing knowledge base in the field of Turkish sentiment analysis.

This corpus, given in Figure 3, is primarily composed of three sources: the TDK Turkish Dictionary [Türk Dil Kurumu, 25], the dictionary of idioms (https://deyimlerimiz.com/), and open-source literary works.

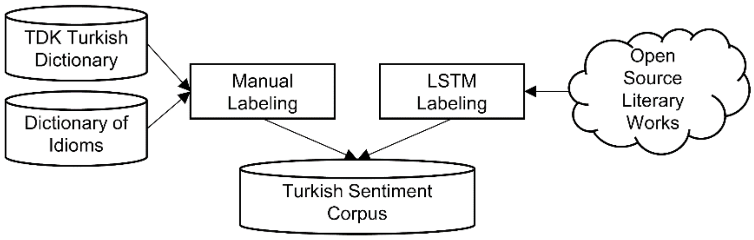

Figure 3: Steps for Creating the Turkish Corpus

The primary data source for the Turkish sentiment corpus is the TDK Turkish Dictionary. Example sentences for the words listed in this dictionary were added to the corpus. An example of a sentiment word and its corresponding source page is provided in Figure 4.A.

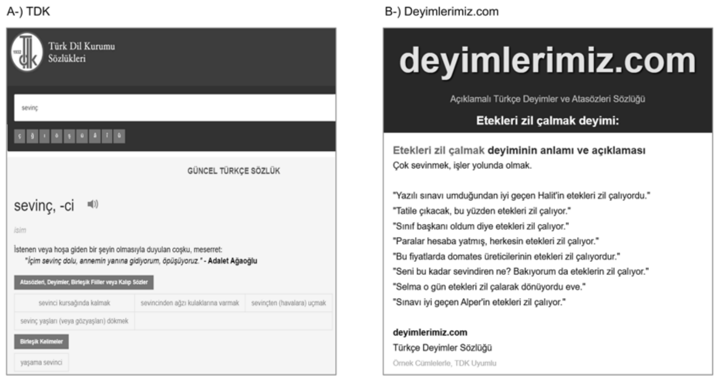

*Figure 4: (a) The representation of the word "sevinç – **joy**" in the TDK Current Turkish Dictionary https://sozluk.gov.tr/; (b) The representation of the idiom "etekleri zil çalmak" on the website https://deyimlerimiz.com/. Both sources are utilized to extract example sentences reflecting sentiment for inclusion in the Turkish sentiment corpus.*

In Figure 4.A, the word the word "sevinç –**joy**" is shown along with its example sentence. Below the example sentence, links to proverbs, idioms, fixed expressions, compound words, and verbs containing the word are provided. In this step, by following both the main page and the links, 34,360 example sentences expressing sentiment were added to the corpus.

Turkish has a rich linguistic structure, with numerous proverbs, idioms, compound verbs, and fixed expressions. Many of these expressions convey emotions. To identify these patterns and label their emotional content, online corpus of Turkish idioms (https://deyimlerimiz.com/) was used as the second data source. This website contains a wide range of idioms along with example sentences. For instance, the idiom "etekleri zil çalmak - **to be on cloud nine**" associated with the word "sevinç – **joy**" is shown in Figure 4.B.

Eight example sentences listed for the relevant idiom are added to the corpus, linking both the idiom itself and the emotion of joy. Similarly, all idioms associated with a specific emotion was scanned from the relevant pages, and 30,000 sentences were added to the corpus. Additionally, idiomatic emotional verbs not found on the website were also considered. For these expressions, Sas's updated corpus [Şaş, 23] was used, and the 707 idiomatic emotional expressions from this corpus were also incorporated into the Turkish sentiment corpus.

During the manual labelling of sentences, Plutchik's Wheel of Emotions, intensifiers and mitigators, non-monotonic logical approaches, emotion-expressing punctuation marks, and morfo-semantic affixes was taken into account. During this phase, the annotation process was guided by Plutchik's Wheel of Emotions, taking into account the primary and secondary emotions, as well as their corresponding intensity levels, as elaborated in Section 3.

In addition to manual labelling, an automatic labelling mechanism was also implemented. For this purpose, a web scraper was designed to download open access literary sources online. These sources are divided into sentences, and subsequently into words, which are processed and passed to the LSTM model for automatic labeling.

The Turkish sentiment corpus, created using the methods described in the paragraphs above, currently consists of 136,000 sentences, with this number steadily increasing[1].

4.2 Methods

This section provides the details of the labeling, vectorization, and LSTM modules of the Duygu-Turk model, under the following subsections.

4.2.1 Labelling Module

The sentences in the Turkish Corpus are labeled in two separate stages: manual and automatic. In the manual labeling, sentences containing sentiment expressions are labelled based on the TDK dictionary. In this step, intensifiers and mitigators, punctuation marks, and morpho-semantic structures, which are not considered in other studies, are taken into account. Additionally, a non-monotonic approach is adopted, and conditional sentiment markers are identified.

Intensifiers and mitigators are crucial elements in the labelling process, as they directly affect sentiment intensity. In Turkish sentence, words such as "çok – very" and "daha – more" function as intensifiers, while "az – less" and "biraz – a little" serve as mitigators. During dataset annotation, sentiment labels were assigned values that reflect sentiment intensity. Sentiment intensity is rated on a scale between 0 – 1 (absence of sentiment - presence of sentiment). Table 1 presents the combinations of intensifiers and mitigators, along with their corresponding coefficients representing sentiment intensities.

Intensifiers - Mitigators and Their Combinations	Sentiment Intensity Coefficient	Intensifiers - Mitigators and Their Combinations	Sentiment Intensity Coefficient
Çok – ***very***	1.2	Haddinden fazla – ***beyond measure***	1.8
Pek – ***quite***	1.2	Son derece – ***extremely***	1.8
Büyük – ***great***	1.2	Derin – ***deeply***	1.8
En – ***most***	1.2	Az – ***little***	0.8
Daha Çok – ***more***	1.4	Biraz – ***a bit***	0.8
Pek Çok – ***quite a lot***	1.4	Küçük – ***small***	0.8
Çok Fazla – ***too much***	1.4	Çok Az – ***very little***	0.6
Pek Fazla – ***quite too much***	1.4	Daha Az – ***less***	0.6
Çok Ama Çok – ***very, very much***	1.4	Pek Az – ***quite little***	0.6
Çok Daha Fazla – ***much more***	1.6	Daha da Az – ***even less***	0.4

[1] Turkish sentiment corpus is available via the following link: https://github.com/rabia-tintin/turkish-plutchik-emotion-dataset.

Çok Çok Fazla – ***extremely much***	1.6	Çok Daha Az – ***much less***	0.4
Daha da Çok – ***even more***	1.6	Azıcık – ***tiny bit***	0.2
Aşırı – ***excessive***	1.8	Minicik – ***very tiny***	0.2

Table 1: Intensifiers/Mitigators and Their Corresponding labels

In recent studies [Karamollaoğlu, 18], [Shehu, 19], [Ahmetoğlu, 20], [Tuzcu, 20] the use of punctuation marks that convey sentiment has been largely overlooked. In Turkish, exclamation marks (!), question marks (?), and exclamation marks within parentheses ((!)) express different sentiments [Türk Dil Kurumu, 24], [Durukoğlu, 18]. While the exclamation mark indicates emotions such as joy, pride, pain, fear, or surprise, the question mark conveys doubt, curiosity, or concern. The parenthesized exclamation mark, on the other hand, represents sarcasm, irony, or contempt. In this study, punctuation marks reflecting sentiment were encoded as distinct characters for proper representation.

Similar to punctuation marks, certain suffixes in Turkish impart emotional meaning to the words they attach to [Akçataş, 21]. This phenomenon, known as morpho-semantic processing. It occurs when a suffix and a word stem, neither of which inherently carry sentiment, combine to create an emotional meaning. For instance, the suffix "-cağız" and the word "çocuk – child" do not individually convey sentiment. However, when combined as "çocukcağız - poor little child", the resulting word expresses pity and compassion. Considering this effect, suffixes such as "-cağız, -ceğiz", which convey pity and compassion, and "-sa, -se," which express desire and expectation, were included in the corpus.

Existing studies [Uyaroğlu Akdeniz, 21], [Aslan, 23], [Polat, 22], [Shehu, 21], [Bozuyla, 24], [Tohma, 23], [Tepecik, 24], [Savci, 23], [Uca, 22] utilize a monotonic logic approach, categorizing emotions into a maximum of three labels: positive, negative, and neutral. However, this approach classifies emotional states as definitive, even when they are not fully determined. As a result, it fails to capture context-dependent emotions and emotional transitions [Boratav, 11]. To address this limitation, non-monotonic logic [Reiter, 80] was adopted. This approach accounts for "conditionally positive" and "conditionally negative" emotions, where initial classifications as positive or negative may be invalidated or altered based on context. A sample of the labeled data using this method is presented in Table 2.

Example Sentence	Semantic Result	Explanation
Söylediklerimi yaparsan ikimiz de mutlu oluruz. (*If you do what I say, both of us will be happy.*)	conditionally positive	*Being happy is conditioned on whether or not the stated actions are carried out. No one has become happy yet*

Yüksek not alırsan dünyalar benim olur. (*In case you get a high grade, the world will be mine.*)	conditionally positive	*The phrase "Dünyalar benim olur" is an expression that signifies happiness. Happiness is conditioned on whether or not a high grade is achieved. No one has experienced happiness yet*.
İşe geç kalırsak patron çok öfkelenir. (*Being late for work will make the boss very angry.*)	conditionally negative	*The boss's anger" is conditional upon whether or not one is late for work. No one has experienced anger ye*t.)
Kendisine iki laf edilse, öfkeden gözleri yuvalarından fırlar. (*Should anyone say a word to him, his eyes would pop out of their sockets from anger.*)	conditionally negative	*The idiom "his eyes would pop out of their sockets" is used in this sentence to mean "to get angry." The feeling of anger is conditional upon whether or not someone says a word to him. As of now, no one has become angry.*

Table 2: Example Sentences for Conditional Positive and Conditional Negative Classes

To ensure consistency and reliability in the manual labeling, three native Turkish speakers with a linguistic background were employed as annotators. They followed a standardized annotation guideline document, which included formal definitions and examples for each of the 16 sentiment categories, along with rules for resolving ambiguity in edge cases. These definitions and decision rules—particularly regarding conditional sentiment expressions, intensifiers, mitigators, punctuation marks, and morpho-semantic cues—are detailed in this section.

The Fleiss' Kappa metric [Fleiss, 71] was used to show the agreement between annotators. This metric produces results that are more reliable when the number of annotators is more than two [Sim, 05]. Fleiss' Kappa score was calculated as 0.81. This shows substantial agreement between the annotators.

In addition to manual labeling, an automatic labeling is also performed. Using a web scraper, open-access literary sources available on the web are downloaded, divided into sentences, and then into words, undergoing preprocessing. Subsequently, the preprocessed data is transferred to the trained LSTM model for labeling.

4.2.2 Vectorization Module

This module aims to generate sentence vectors from the dataset. Throughout the process, word vectors are obtained by combining the UniGram method and rule-based features.

Various studies in the literature [Kim, 16], [Ali, 23] have demonstrated that the UniGram method offers notable advantages for modeling morphologically rich languages at the character level. In particular, it effectively captures semantic similarities between word roots and their derivatives, operates independently of language-specific rules, features a simpler structure compared to more complex models, and requires fewer computational resources [Gerz, 18]. Moreover, preliminary studies have shown that the UniGram method is more effective in adapting to Turkish's agglutinative structure and representing fixed expressions than popular word embedding techniques like FastText and GloVe. Due to these advantages, the UniGram was chosen for the character-level vector representation of words.

In the vectorization module, each character of the word root is converted into vectors using UniGram. During this phase, the frequency of the usage of Turkish letters [Güneş, 18] is also considered. The vector values of characters and sentiment-indicating punctuation marks, based on their usage frequency, are presented in Table 3.

Char.	**Vector Value**	**Char.**	**Vector Value**	**Char.**	**Vector Value**	**Char.**	**Vector Value**	**Char.**	**Vector Value**
a	0.01	**t**	0.09	**ü**	0.17	**v**	0.25	**! > x**	0.33
e	0.02	**ı**	0.10	**ş**	0.18	**ğ**	0.26	**? > q**	0.34
k	0.03	**s**	0.11	**z**	0.19	**ö**	0.27	**(!) > w**	0.35
ı	0.04	**u**	0.12	**g**	0.20	**f**	0.28		
l	0.05	**y**	0.13	**ç**	0.21	**j**	0.29		
m	0.06	**d**	0.14	**p**	0.22	**â**	0.30		
r	0.07	**o**	0.15	**h**	0.23	**î**	0.31		
n	0.08	**b**	0.16	**c**	0.24	**û**	0.32		

Table 3: Vector Values Based on the Frequency of Use of Turkish Letters

The vector values in Table 3 were determined inversely proportional to the frequency of letter usage in Turkish. More frequently used letters have lower vector values. This minimizes large weight changes during the model's parameter updates and contributes to a more stable and efficient optimization process. Additionally, it reduces training time and computational costs by ensuring consistent calculation of gradients. Furthermore, representing less frequent letters with higher values aids the model in better learning rare characters, thereby enhancing overall performance.

During the data pre-processing phase, inflectional suffixes were removed through stemming, which reduces the length of words. The longest word in a Turkish corpus, after removing suffixes, consists of 23 characters [Dalkılıç, 03]. Based on this, the dimension of the UniGram vectors was set to 20. If a word is shorter than 20 characters, padding is applied for the remaining vector dimensions. In rare cases where the word length exceeds 20 characters, truncation is applied.

Another structure used in the vectorization module is the rule base. The rule base reflects the attributes that could contribute to the sentiment analysis of a word into the vector. These attributes include word types (noun, adjective, adverb, pronoun, etc.), suffix types (conditional, negation, wish, etc.), and punctuation marks. A 20-

dimensional space was allocated in the vector for these attributes. The corresponding structure is presented in Table 4.

Index	Attribute	Index	Attribute
20	İsim - ***Noun***	29	Soru İşareti ? – ***Question Mark***
21	Sıfat – ***Adjective***	30	Yay Ayraçlı Ünlem İşareti (!) - ***Interrobang***
22	Zarf – ***Adverb***	31	Yoksunluk Eki (-sız, -siz, -suz, -süz) - ***Suffix of Lack***
23	Ünlem – ***Interjection***	32	Olumsuzluk Eki (-ma, -me) - ***Negation Suffix***
24	Yüklem – ***Predicate***	33	Olumsuz İfade (değil, yok) - ***Negative Expression***
25	Fiilimsi – ***Varbal Noun***	34	Koşul Eki (-sa, -se) - ***Conditional Suffix***
26	Zamir – ***Pronoun***	35	Dilek Eki (-sa, -se) - ***Optative Suffix***
27	Eş Sesli Kelime – ***Homophone***	36	Kuvvetlendirici İfadeler - ***Intensifiers***
28	Ünlem İşareti ! – ***Exclamation Mark***	37	Zayıflatıcı İfadeler - ***Mitigators***
29	Soru İşareti ? – ***Question Mark***	38	Küçültme Eki (-cağız, -ceğiz) - ***Diminutive Suffix***
30	Yay Ayraçlı Ünlem İşareti (!) - ***Interrobang***	39	Karşıtlık Bağlaçları (ama, ancak, fakat, lakin, lâkin, rağmen) - ***Contrastive Conjunctions (e.g., but, however, although, yet, on the contrary)***

Table 4: Attributes of the rule base and their indices within the word vector

Rule-based features enhance the accuracy of sentiment analysis by modeling the grammatical and semantic structure of words. Features related to word types (noun, adjective, adverb, exclamation, verb, verbal noun, pronoun) was added to the vector to aid in understanding sentence structure. The homophone feature is used to prevent incorrect root detection and to increase attention on these words.

The suffix of lack (-sız, -siz, -suz, -süz), the negation suffix (-ma, -me), and negation expressions (değil, yok) are placed in the range of indices 31 to 33 of the rule-based vector, as they can alter the sentiment of the sentence.

The conditional suffix (-sa, -se) is critical for detecting conditional sentiment states within the non-monotonic logic. For example, in the sentence "Hayallerimi gerçekleştiremezsem çok üzülürüm - **If I can't make my dreams come true, I will be very sad**", the state of being sad is dependent on a condition. The person has not yet experienced sadness and the state of sadness may change depending on whether the dreams come true. Therefore, classifying the sentence as "negative" would not be accurate. Since "being sad" is a negative sentiment, but it is conditional, it is more accurate to classify it as "conditional-negative." To capture such conditional expressions, words with conditional suffixes are represented in the rule-based vector approach.

In Turkish, the conditional suffix and the optative suffix are the same (-sa, -se). For example, in the sentence "Yarın bayram olsa - **I wish tomorrow were a holiday**", the suffix -sa expresses emotions like desire or expectation. However, in "Yarın bayram olsa coşku ile kutlardık - **If tomorrow were a holiday, we would celebrate with enthusiasm**", the same suffix indicates a condition. The distinction between these uses depends on the position of the suffix within the sentence. In Turkish, when -sa, -se is used as an optative suffix, it typically appears at the end of the sentence. This rule was incorporated into the rule-based approach, where the suffix is assigned to the relevant index based on its position in the sentence (34 for condition, 35 for optative).

Since intensifiers and mitigators modify sentiment intensity, they are also represented in the rule-based vector. Similarly, the diminutive suffix (-cağız, -ceğiz), which conveys a sense of pity, is included in the rule-based vector.

Detecting transitions between opposing sentiments is crucial for improving the accuracy. For example, in the sentence "Beni üzmesine rağmen halen onu seviyorum - **Although it upsets me, I still love him/her**", there is a transition from sadness to joy through a contrastive conjunction. In this case, both emotions coexist within the sentence. To identify such sentiment transitions, contrastive conjunctions (e.g., ama, ancak, fakat, lakin, lâkin, rağmen) were incorporated as a feature in the rule-based vector.

In the vectorization module, word vectors are formed by combining a 20-dimensional UniGram vector and a 20-dimensional rule-based vector for each word. Subsequently, the word vectors of all words in a sentence are merged to create the sentence vector. Studies indicate that the average number of words in Turkish sentences ranges between 10 and 15 [Dalkılıç, 03]. To account for exceptional cases where sentence length may exceed this range, the maximum number of words per sentence has been set to 25. Based on this, sentence vectors are structured to be 1000-dimensional. Padding is applied to fill empty vector spaces, while truncation is used for sentences exceeding 25 words.

In summary, the UniGram Word Vector captures sequential characteristics in a word's root and affix structures at the character level, aiding LSTM in learning sequential information and morphological regularities. Meanwhile, the Rule-Based Approach Vector highlights linguistically relevant features for sentiment analysis, enabling the model to grasp grammatical details. The combination of these two vectors allows the LSTM model to learn both the statistical regularities and rule-based structure of the language. As a result, the model can make more accurate and consistent predictions based on semantic and contextual relationships in the language.

4.2.3 LSTM Module

The Long Short-Term Memory (LSTM) networks are an extension of recurrent neural networks (RNNs) designed for modeling sequential data [Hochreiter, 97]. Traditional RNN models face significant challenges, such as the vanishing gradient problem, in maintaining long-term contextual information. As illustrated in Figure 5, LSTM addresses this issue through specialized mechanisms [Gers, 00].

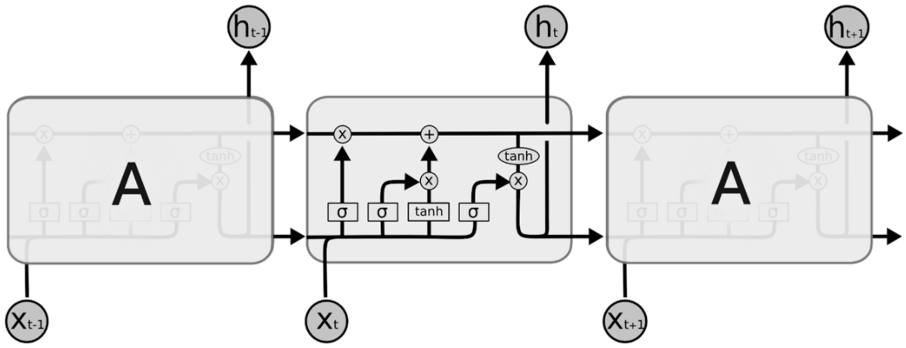

Figure 5: LSTM architecture

The LSTM operates using three types of gates: the forget gate (f_t) determines which information should be discarded, the input gate (i_t) controls which information should be added, and the output gate (o_t) decides which information should be passed to the next state. The weight matrices regulate the extent to which each gate is affected. The equations for these gates are provided below, where W represents the weight matrix, b denotes the bias vector, and σ is the sigmoid activation function [Yalman, 22].

Forget Gate: $f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f)$ (1)

Input Gate: $i_t = \sigma(W_{xi}\,x_t + W_{hi}\,h_{t-1} + W_{ci}\,c_{t-1} + b_i)$ (2)

Cell State Update: $c_t = f_t\,c_{t-1} + i_t\,tanh(W_{xc}\,x_t + W_{hc}\,h_{t-1} + b_c)$ (3)

Output Gate: $o_t = \sigma(W_{xo}\,x_t + W_{ho}\,h_{t-1} + W_{co}\,c_t + b_o)$ (4)

Hidden State Update: $h_t = o_t\,tanh(c_t)$ (5)

Through these control mechanisms, LSTM effectively retains past information, allowing it to capture long-term dependencies in the input more efficiently. This makes it a powerful method for sentiment analysis by preserving sequential context in textual data. The ability to model and recall dependencies provides a significant advantage in languages like Turkish, which has a complex morphological structure, an agglutinative nature, and relies heavily on contextual information.

In this study, the LSTM model is trained with 1000-dimensional sentence vectors, which combine UniGram and rule-based features transformed into word embeddings. The LSTM layers then process these vectors to build the sentiment classification model. In order to establish the most effective model, hyperparameters of the LSTM model (both training-related and architectural parameters) was optimized. In this step, the evolutionary search [Bäck, 93] was employed. Computational efficiency was also considered during optimization. The optimal values for these parameters are presented in Table 5.

Hyperparameters	Optimum Values	**Hyperparameters**	Optimum Values
Batch Size	256	Layer 3	Dropout
Learning Rate	0.01	Layer 4	Output
Dropout Rate	0.3	Activation Function 1	ReLU
Epoch	1000	Activation Function 2	ReLU
Weight Initialization Methods	RELU UNIFORM	Number of Neurons in Hidden Layer 1	2048
Optimizer	Adam	Number of Neurons in Hidden Layer 2	512
Layer 1	Dense	CPU-Total time	05:41 h
Layer 2	LSTM	Model Size	54.530 KB

Table 5: LSTM Model Hyperparameters and Their Optimum Values

In the optimized model, the ReLU activation function was used in both the Dense and LSTM layers. For the output layer, the softmax activation function was chosen. Softmax is commonly used in multi-class classification as it transforms output values into a probability distribution [Bridle, 89], [Bridle, 90]. The class labels and their corresponding values are presented in Table 6.

Class Label	**Value of Labels**
Morphological Result	{Positive, Negative, Neutral, Conditional Positive, Conditional Negative}
Sentiment Result	
Joy	{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8}
Trust	{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8}
Fear	{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8}
Surprise	{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8}
Sadness	{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8}
Disgust	{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8}
Anger	{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8}
Anticipation	{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8}

Table 6: LSTM Class Labels and Value of Labels

According to this structure, the first class label determines the morphological outcome (Positive, Negative, Neutral, Conditional Positive, Conditional Negative). The second class label, the semantic outcome, is used to identify the primary emotions in Plutchik's wheel of emotions. In addition, the intensity level of the primary emotion can be determined by using the coefficients appropriate to the intensity change in the emotion wheel. Secondary emotions are also detected by considering the intersection set of primary emotions and their intensities.

To mitigate the potential impact of class imbalance in the training and test datasets—caused by the large number of classes—a weighted classification approach was employed. This method ensures that the model does not become biased toward more frequent classes. The weight for each class was determined as in equation 6.

$$W_i = \frac{1}{Cf_i} \quad (6)$$

In this equation Wi demonstrates the weight for a given class i and Cf_i is the frequency of class i in the dataset. As the number of classes increases or the imbalance becomes more prominent, the weights become more important in training the model.

5 Case Study

In order to explain the decision-making process of the Duygu-Turk Model, this section presents a scenario based on a sample sentence. In this scenario, the example sentence first goes through the preprocessing step and then is given to the vectorization module. Here, it is converted into a sentence vector and then labeled by the LSTM model.

In the pre-processing step, all punctuation marks in the sentence except exclamation marks, question marks, and exclamation marks within parentheses are removed; all letters are converted to lower case; and the stems of all words are identified. The sample sentence and data pre-processing step are presented in Table 7.

Input	Hayallerimizden vazgeçmezsek her şeyin çok güzel olacağına inanıyorum. (*I believe that if we don't give up on our dreams, everything will be much beautiful*)
Data Pre-processing	hayal vazgeç her şey çok güzel ol inan (*dream give up every thing much beautiful be beleive*)

Table 7: Example sentence and result of preprocessing

After pre-processing, the sentence is split into words. A 40-dimensional vector is constructed for each word. The vector segment indexed from 0 to 19 (Table 3) is allocated to UniGram features, while the segment indexed from 20 to 39 (Table 4) is assigned to Rule-Based word features. Each word is analyzed in turn. Unigram and rule based word vectors resulting from this analysis are presented in Table 8.

Word	UniGram (20 dimension)	Rule Based (20 dimension)
hayal *(Dream)*	0.23,0.01,0.13,0.01,0.05,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
vazgeç (*give up)*	0.25,0.01,0.19,0.20,0.02,0.21,0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,1,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0
her (*every)*	0.23,0.02,0.07,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
şey (*thing)*	0.18,0.02,0.13,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
çok (much)	0.21,0.15,0.03,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0
güzel (beautiful)	0.20,0.17,0.19,0.02,0.05,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
ol (be)	0.15,0.05,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0
inan (believe)	0.04,0.08,0.01,0.08,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0	0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0

Table 8: UnigGram and Rule Based vector representations

The analysis of semantically critical words in the sentence is performed as follows:

The second word in the sentence, "vazgeç – give up", is converted into UniGram features based on the frequency values provided in Table 3, resulting in the vector [0.25, 0.01, 0.19, 0.20, 0.02, 0.21]. Padding is applied to the remaining dimensions of the vector. From a linguistic feature perspective, "vazgeç – give up" functions as the predicate of the subordinate clause, contains a negation suffix, and includes a conditional suffix. Therefore, in the rule-based word vector (Table 4), the 24th, 32nd, and 34th index values are set to "1".

The fifth word in the sentence, "çok – much", is converted into UniGram features based on the frequency distribution of Turkish characters provided in Table 3, resulting in the vector [0.21, 0.15, 0.03]. Padding is applied to fill the remaining dimensions of the UniGram word vector. Subsequently, the word is analyzed using the rule-based feature set defined in Table 4. Since "çok – much" functions as an adjective and serves as an intensifier, the 21st and 36th indices in the rule-based vector are set to "1". The resulting rule-based feature vector for this word is constructed as [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0].

After generating both UniGram and rule-based vectors for each word in the sentence, these vectors are concatenated to form the final word representation. The resulting word vectors for the example sentence are presented in Table 9.

Word	Word Vector (40 dimension)
hayal *(Dream)*	0.23,0.01,0.13,0.01,0.05,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
vazgeç (*give up)*	0.25,0.01,0.19,0.20,0.02,0.21,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,0,1,0,0,0,0,0

her (*every)*	0.23,0.02,0.07,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0
şey (*thing)*	0.18,0.02,0.13,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0
çok (much)	0.21,0.15,0.03,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,1,0,0,0
güzel (beautiful)	0.20,0.17,0.19,0.02,0.05,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0, 0,0,0,0,0,0,0,0,0,0,0,0
ol (be)	0.15,0.05,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0, 0,0,0,0,0,0,0
inan (believe)	0.04,0.08,0.01,0.08,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0 ,0,0,0,0,0,0,0,0,0,0

Table 9: Generation of word vectors.

When a sentence consists of sub-sentences containing more than one judgment (comma, semicolon connected sentences), the word vectors created for each sub-sentence are merged and transferred to the sentence vector. For a structure consisting of a single sentence as in the example, padding is applied for the remaining indices of the sentence vector.

Once the sentence vector is generated, it is fed into the LSTM model. The model produces two main results: semantic and morphological. In the morphological result, according to the non-monotonic approach, the sentence is assigned to one of the positive, negative, neutral, conditional positive, conditional negative classes. In the semantic result, the sentence is labeled according to the eight basic emotions in Plutchnik's emotion wheel. At this stage, the model calculates an intensity for each primary emotion. The emotion with the highest intensity is set as the primary emotion label. When multiple primary emotions exhibit high intensity values, the corresponding secondary emotions are inferred based on the intersections of those primary emotions on Plutchik's wheel of emotions.

The semantic and morphological outputs for the example sentence are shown in Figure 6.

Semantic Output	Morphological Output
-> Joy > 1,2 > 1,0000000000 > Joy -> Anticipation > 1,2 > 1,0000000000 > Anticipation -> Trust > 0,0 > 1,0000000000 > -> Fear > 0,0 > 1,0000000000 > -> Surprise > 0,0 > 1,0000000000 > -> Sadness > 0,0 > 1,0000000000 > -> Disgust > 0,0 > 1,0000000000 > -> Anger > 0,0 > 1,0000000000 > --------------------------------- SECONDARY EMOTION : OPTIMISM ---------------------------------	-> Conditional-Positive: 1.0 -> Positive: 2.4117177699345405E-10 -> Conditional-Negative: 1.3612989208100501E-11 -> Neutral: 6.297161576906962E-12 -> Negative: 1.404558029128103E-13

Figure 6: The semantic and morphological outputs

From a structural perspective, it is observed that the LSTM model effectively learns conditional patterns. Accordingly, by taking into account the conditional suffix indicator in the word vector of the example sentence (i.e., the 34th index), the model assigns the label conditional positive with the highest confidence.

From a semantic perspective, the two primary emotions with the highest weights are joy and anticipation. Therefore, the intersections of these primary emotions, as illustrated in Figure 7, was taken into consideration.

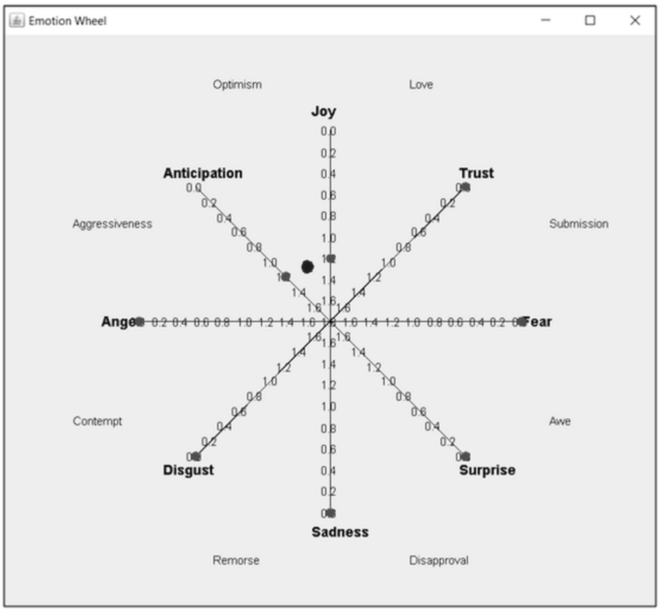

Figure 7: Primary and secondary emotions

According to Figure 7, when the emotions of anticipation and joy intersect, the emotion of optimism emerges as a secondary emotion. Based on this analysis, when an example sentence is labeled using LSTM, the morphological output is conditionally positive. In terms of semantic results, the primary emotion labels are Joy and Expectation, while the secondary emotion label is Optimism.

6 Findings

Duygu-Turk produces two main outputs. In the morphological output, by adopting a non-monotonic logic approach, not only positive and negative labels are identified, but also conditional positive, conditional negative and neutral labels. After this preliminary classification, the eight primary and eight secondary emotions and their intensity levels in Plutchik's wheel of emorions are determined.

The sentiment corpus detailed in section 4.1 was used for training, testing and validation of the model. Out of this corpus, 100,000 sentences were selected by stratified sampling and used for training and testing. In this step, fivefold cross validation scheme was used. A set of 36,000 sentences not used in the training and testing step was used for model validation. The morphological and semantic classification performance of the model was evaluated based on accuracy, precision, recall and F1 score metrics.

The classification performances according to the non-monotonic logic approach are presented in Table 10.

Performance / Class	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Negative	0.9435	0.9356	0.9454	0.9405
Positive	0.9481	0.9041	0.8929	0.8985
Neutral	0.9729	0.9038	0.8910	0.8974
Conditional Positive	0.9965	0.9731	0.9690	0.9710
Conditional Negative	0.9957	0.9707	0.9752	0.9729
Macro Average	0.9713	0.9374	0.9347	0.9360
Weighted Average	0.9558	0.9283	0.9284	0.9283

Table 10: Classification performance values based on non-monotonic logic approach

For the classification based on non-monotonic logic, macro and weighted average F1 scores were 93% and 92%, respectively. It is observed that the highest classification performance of Govde-Turk is achieved in the categories of "conditionally positive" and "conditionally negative." This indicates that the representation of the conditional suffix ("-sa, -se") in the rule-based vector yields effective results, enabling the model to accurately capture emotions that are dependent on conditions.

The classification performance for the primary emotions in the emotion wheel are presented in Table 11.

Perf. / Class	Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger	Anticipation	Macro AVG
Acc.	0.95	0.97	0.96	0.97	0.95	0.96	0.96	0.96	0.96
Prec.	0.87	0.94	0.89	0.92	0.85	0.90	0.94	0.90	0.90
Rec.	0.83	0.90	0.92	0.92	0.85	0.88	0.93	0.92	0.89
F1	0.85	0.92	0.90	0.92	0.85	0.89	0.93	0.91	0.90

Table 11: Classification performance of Duygu-Turk for primary emotions

The macro average values obtained in the classification of primary emotions were 0.96, 0.90, 0.89 and 0.90 for accuracy, precision, recall and F1, respectively.

The Duygu-Turk model was also compared with BERT, distilled BERT and ELECTRA models, which are known for their successful performance on the Turkish language. These transformer-based models can only classify emotions as positive, negative and neutral. Therefore, the performance of the competing models on the Turkish sentiment corpus without pre-training is quite low. In the preliminary study, the performance metrics for BERT were 0.0159 accuracy, 0.8 precision, 0.0159 recall and 0.0076 F1 score. The competing models were then put through a pre-training phase and performance comparison was performed. In the first step, only positive and negative classification was performed. The results of this comparison are presented in Table 12.

	Duygu-Turk	BERT	ELECTRA	Distilled BERT
Accuracy	***0.9977***	0.9445	0.9652	0.8429
Precision	***0.9985***	0.9692	0.9768	0.9312
Recall	***0.9989***	0.9445	0.9652	0.8429
F1 Score	***0.9987***	0.9520	0.9685	0.8763

Table 12: Duygu-Turk's performance comparison with current competitors (positive, negative)

The results reveal that Duygu-Turk outperforms the competing models in positive, negative and neutral classification. Among the competing models, the highest value in terms of F1 score belongs to the ELECTRA (0.96). However, Duygu-Turk increased this value to 0.99 and became the most successful model.

In the second step, all models were compared for the classification of the eight primary emotions in the emotion wheel. The results are given in Table 13.

	Sadness (%)				**Disgust (%)**				**Anger (%)**			
	DT	B	ELC	DB	**DT**	B	ELC	DB	**DT**	B	ELC	DB
Acc	***95***	8	18	10	***96***	64	21	48	***96***	83	26	21
Prec	***85***	84	88	87	***90***	88	86	86	***94***	90	81	82
Rec	***85***	8	18	10	***88***	64	21	48	***93***	83	26	21
F1	***85***	10	17	3	***89***	70	18	56	***90***	85	17	14

	Joy (%)				**Trust (%)**				**Fear (%)**			
	DT	B	ELC	DB	**DT**	B	ELC	DB	**DT**	B	ELC	DB
Acc	***95***	81	5	48	***97***	81	74	38	***96***	71	31	53
Prec	***87***	92	1	90	***94***	93	92	90	***89***	90	87	88
Rec	***83***	81	5	48	***90***	81	74	38	***92***	71	31	53
F1	***85***	85	1	59	***92***	85	79	47	***90***	76	33	62

	Anticipation (%)				**Surprise (%)**			
	DT	B	ELC	DB	**DT**	B	ELC	DB
Acc	***96***	60	70	65	***97***	85	85	66
Prec	***90***	88	89	88	***92***	92	93	89
Rec	***92***	60	70	65	***92***	85	85	66
F1	***91***	66	76	72	***92***	87	87	7

Table 13. Performance comparison of Duygu-Turk (DT) and current competitors (B: BERT, ELC:ELECTRA, DB: Distilled BERT) in the classification of eight primary emotions

In the evaluations conducted for the classification of eight primary emotions, Duygu-Turk achieved the highest score in each category. These results clearly demonstrate the model's effectiveness in sentiment analysis tasks and indicate that the proposed method offers a performance that surpasses existing models.

7 Conclusion

This study introduced Duygu-Turk, a sentiment analysis model specifically developed for the Turkish language, addressing key limitations in prior research by incorporating emotion intensity, secondary emotions, and linguistic features unique to Turkish. Built upon Plutchik's wheel of emotions and enriched with non-monotonic logic, the model enables fine-grained classification across eight primary and eight secondary sentiment categories, including conditionally expressed emotions, which are often overlooked in existing approaches.

Experimental results demonstrate the superiority of Duygu-Turk over established transformer-based models such as BERT, distilled BERT, and ELECTRA. While these models are constrained to three-class sentiment categorization and exhibit limited performance on Turkish without extensive pretraining, Duygu-Turk consistently outperforms them across all evaluated tasks. Notably, the model achieved macro and weighted average F1 scores of 93% and 92% respectively for conditional sentiment detection, and reached an F1 score of 0.99 in standard polarity classification—surpassing the closest competing model, ELECTRA.

Large-scale pre-trained models such as BERTurk provide strong linguistic representations for Turkish. However, our empirical findings indicate that, without domain-specific fine-tuning, these models underperform in nuanced sentiment classification tasks. Moreover, the fine-tuning process itself can be resource-intensive in terms of both computational cost and time. In contrast, our LSTM-based approach—specifically designed to integrate linguistically informed features—offers greater interpretability and aligns well with the fine-grained emotion taxonomy employed in this study.

Furthermore, Duygu-Turk proved highly effective in classifying the eight primary emotions, with superior performance metrics in all categories. These findings confirm that incorporating morpho-semantic elements, idiomatic structures, and context-sensitive logic significantly enhances sentiment classification in Turkish. By offering a linguistically grounded, computationally efficient, and empirically validated approach, Duygu-Turk establishes a new benchmark for sentiment analysis in low-resource, morphologically rich languages.

8 Limitations and Future Work

Due to the lack of sufficiently comprehensive and large-scale resources for Turkish sentiment analysis, a manually annotated corpus had to be constructed initially. However, this limitation has been largely addressed through the integration of an automatic labeling mechanism. By leveraging large-scale literary corpora and idiomatic resources, the system now supports semi-supervised expansion and automatic vector assignment, enhancing its scalability and reducing manual dependency. Furthermore, while the model has already been tested on a large corpus including real-world texts, its performance on domains that exhibit highly informal, abbreviated, or evolving language—such as social media content—remains an area for further exploration.

As part of future work, we aim to transform the current rule-based LSTM architecture into a hybrid framework that incorporates transformer-based models, such as BERT or RoBERTa. This hybridization is expected to improve the model's contextual understanding, handle complex syntactic structures more effectively, and increase robustness against informal or unseen linguistic patterns. Moreover, the integration of contextual embeddings will allow for more nuanced sentiment detection, especially in cases involving sarcasm, idiomatic expressions, or polysemy.

References

[Ahmetoğlu, 20] Ahmetoğlu, H., Daş, R.: Türkçe Otel Yorumlarıyla Eğitilen Kelime Vektörü Modellerinin Duygu Analizi ile İncelenmesi. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 24(2), 2020, 455-463.

[Akçataş, 21] Akçataş, A., Varol, Ç.: Türkiye Türkçesinde İsimden İsim Yapan Eklerin Morfo-Semantik İşlevleri. Kesit Akademi Dergisi, 7(26), 2021, 410-427.

[Alawi, 24] Alawi, A. B., Bozkurt, F.: A Hybrid Machine Learning Model for Sentiment Analysis and Satisfaction Assessment with Turkish Universities Using Twitter Data. Decision Analytics Journal, 11, 2024, 100473.

[Alkan, 23] Alkan, B. B., Karakus, L., Direkci, B.: Generating A Sentiment Dictionary in R And Dictionary-Based Sentiment Analysis in Turkish Texts. Digital Scholarship in the Humanities, 38(2), 2023, 461-476.

[Ali, 23] Ali, S., Murat, A.: Improved Character-Based Neural Network for POS Tagging on Morphologically Rich Languages. Journal of Information Processing Systems, 19(3), 2023, 355-369.

[Alonso, 21] Alonso, M. A., Vilares, D., Gómez-Rodríguez, C., Vilares, J.: Sentiment Analysis for Fake News Detection. Electronics, 10(11), 2021, 1348.

[Aslan, 23] Aslan, S.: Doğal Dil İşleme Teknikleri Kullanarak E-Ticaret Kullanıcı İncelemelerinde Özellik Tabanlı Duygu Analizi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 35(2), 2023, 875-882.

[Bäck, 93] Bäck, T., Schwefel, H. P.: An Overview of Evolutionary Algorithms for Parameter Optimization. Evolutionary Computation, 1(1), 1993, 1-46.

[Balli, 22] Balli, C., Guzel, M. S., Bostanci, E., Mishra, A.: Sentimental Analysis of Twitter Users from Turkish Content with Natural Language Processing. Computational Intelligence and Neuroscience, 2022(1), 2022, 2455160.

[Bozuyla, 24] Bozuyla, M.: Sentiment Analysis of Turkish Drug Reviews with Bidirectional Encoder Representations from Transformers. ACM Transactions on Asian and Low-Resource Language Information Processing, 23(1), 2024, 1-17.

[Boratav, 11] Boratav, H. B., Sunar, D., Ataca, B.: Duyguları Sergileme Kuralları ve Bağlamsal Belirleyicileri: Türkiye'de Üniversite Öğrencileri ile Yapılan Bir Araştırma. Türk Psikoloji Dergisi, 26(68), 2011, 90-101.

[Bridle, 89] Bridle, J.: Training Stochastic Model Recognition Algorithms as Networks Can Lead to Maximum Mutual Information Estimation of Parameters. Advances In Neural Information Processing Systems, 2, 1989.

[Bridle, 90] Bridle, J. S.: Probabilistic Interpretation of Feedforward Classification Network Outputs, With Relationships to Statistical Pattern Recognition. In Neurocomputing: Algorithms, Architectures and Applications, Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, 227-236.

[Çetin, 18] Çetin, F. S., Eryiğit, G.: Türkçe Hedef Tabanlı Duygu Analizi için Alt Görevlerin İncelenmesi-Hedef Terim, Hedef Kategori ve Duygu Sınıfı Belirleme. Bilişim Teknolojileri Dergisi, 11(1), 2018, 43-56.

[Dalkılıç, 03] Dalkılıç, G., Çebi, Y.: Türkçe Külliyat Oluşturulması ve Türkçe Metinlerde Kullanılan Kelimelerin Uzunluk Dağılımlarının Belirlenmesi. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 5(1), 2003, 1-7.

[Demirtas, 13] Demirtas, E., Pechenizkiy, M.: Cross-Lingual Polarity Detection with Machine Translation. In Proceedings of The Second International Workshop on Issues of Sentiment Discovery and Opinion Mining, 2013, 1-8.

[Durukoğlu, 18] Durukoğlu, S., Büyükelçi, B.: Türk Dilinde Noktalama İşaretlerinin Kullanımı ve Tarihi Gelişimi. Akra Kültür Sanat ve Edebiyat Dergisi, 6(14), 2018, 11-26.

[Fleiss, 71] Fleiss, J. L.: Measuring Nominal Scale Agreement Among many Raters. Psychological bulletin, 76(5), 1971, 378.

[Gerz, 18] Gerz, D., Vulić, I., Ponti, E., Naradowsky, J., Reichart, R., Korhonen, A.: Language Modeling for Morphologically Rich Languages: Character-Aware Modeling for Word-Level Prediction. Transactions of the Association for Computational Linguistics, 6, 2018, 451-465.

[Gers, 00] Gers, F. A., Schmidhuber, J., Cummins, F.: Learning to Forget: Continual Prediction with LSTM. Neural Computation, 12(10), 2000, 2451-2471.

[Güven, 19] Güven, Z. A., Diri, B., Çakaloğlu, T.: Comparison Method for Emotion Detection of Twitter Users. In 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), IEEE, 2019, 1-5.

[Güven, 21] Güven, Z. A.: Comparison of BERT Models and Machine Learning Methods for Sentiment Analysis on Turkish Tweets. In 2021 6th International Conference on Computer Science and Engineering (UBMK), IEEE, 2021, 98-101.

[Güneş, 18] Güneş, F., Işık, A. D.: Türkçede Sık Kullanılan Harfler ve Öğretilmesi. The Journal of Limitless Education and Research, 3(1), 2018, 1-26.

[Hochreiter, 97] Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. Neural Computation, 9(8), 1997, 1735-1780.

[Karamoğlaoğlu, 18] Karamollaoğlu, H., Doğru, İ. A., Dörterler, M., Utku, A., Yıldız, O.: Sentiment Analysis on Turkish Social Media Shares Through Lexicon Based Approach. In 2018 3rd International Conference on Computer Science and Engineering (UBMK), IEEE, September 2018, 45-49.

[Kaya, 24] Kaya, Y. B., Tantuğ, A. C.: BERT2D: Two Dimensional Positional Embeddings for Efficient Turkish NLP. IEEE Access, vol. 12, 2024, 77429-77441.

[Kim, 16] Kim, Y., Jernite, Y., Sontag, D., Rush, A.: Character-Aware Neural Language Models. In Proceedings of The AAAI Conference on Artificial Intelligence, 30(1), March 2016.

[Kumaş, 21] Kumaş, E.: Türkçe Twitter Verilerinden Duygu Analizi Yapılırken Sınıflandırıcıların Karşılaştırılması. Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi, 2(2), 2021, 1-5.

[Ozturk, 22] Ozturk, O., Ozcan, A.: Sentiment Analysis in Turkish Using Transformer-Based Deep Learning Models. In The International Conference on Artificial Intelligence and Applied Mathematics in Engineering, Cham: Springer International Publishing, 2022, 1-15.

[Plutchik, 23] R. Plutchik, Plutchik's Wheel of Emotions - Google Scholar. Accessed: Aug. 07, 2023. [Online]. Available.

[Polat, 22] Polat, H., Ağca, Y.: Tripadvisor Kullanıcılarının Türkçe ve İngilizce Yorumları Kapsamında Duygu Analizi Yöntemlerinin Karşılaştırmalı Analizi. Abant Sosyal Bilimler Dergisi, 22(2), 2022, 901-916.

[Reiter, 80] Reiter, R.: A Logic for Default Reasoning. Artificial Intelligence, 13(1-2), 1980, 81-132.

[Sağlam, 19] Sağlam, F.: Otomatik Duygu Sözlüğü Geliştirilmesi ve Haberlerin Duygu Analizi, Doctoral Dissertation, Hacettepe University, Institute of Science, Ankara, 2019.

[Savci, 23] Savci, P., Das, B.: Prediction of The Customers' Interests Using Sentiment Analysis in E-Commerce Data for Comparison of Arabic, English, and Turkish Languages. Journal of King Saud University-Computer and Information Sciences, 35(3), 2023, 227-237.

[Şaş, 23] Şaş, A. K.: Türkiye Türkçesinde Deyimleşmiş Birleşik Duygu Fiilleri. Korkut Ata Türkiyat Araştırmaları Dergisi, no. 10, 2023, 602-631.

[Shehu, 19] Shehu, H. A., Tokat, S.: A Hybrid Approach for The Sentiment Analysis of Turkish Twitter Data. In The International Conference on Artificial Intelligence and Applied Mathematics in Engineering, Cham: Springer International Publishing, April 2019, 182-190.

[Shehu, 21] Shehu, H. A., Sharif, M. H., Sharif, M. H. U., Datta, R., Tokat, S., Uyaver, S., ... Ramadan, R. A.: Deep Sentiment Analysis: A Case Study on Stemmed Turkish Twitter Data. IEEE Access, 9, 2021, 56836-56854.

[Sim, 05] Sim, J., Wright, C. C.: The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. Physical therapy, 85(3), 2005, 257-268.

[Tepecik, 24] Tepecik, A., Demir, E.: Emotion Detection with Pre-Trained Language Models BERT and ELECTRA: Analysis of Turkish Data. Intelligent Methods in Engineering Sciences, 3(1), 2024, 7-12.

[Tohma, 23] Tohma, K., Okur, H. I., Kutlu, Y., Sertbas, A.: Sentiment Analysis in Turkish Question Answering Systems: An Application of Human-Robot Interaction. IEEE Access, vol. 11, 2023, 66522-66534.

[Tocoglu, 18] Tocoglu, M. A., Alpkocak, A.: TREMO: A Dataset for Emotion Analysis in Turkish. Journal of Information Science, 44(6), 2018, 848-860.

[Türk Dil Kurumu, 24] Türk Dil Kurumu: Noktalama İşaretleri (Açıklamalar), 2024, https://tdk.gov.tr/icerik/yazim-kurallari/noktalama-isaretleri-aciklamalar/

[Türk Dil Kurumu, 25] Türk Dil Kurumu: Türk Dil Kurumu Sözlükleri, 2025, https://sozluk.gov.tr/

[Tuzcu, 20] Tuzcu, S.: Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması. Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi, 1(2), 2020, 1-5.

[Uca, 22] Uca, E., Yılmazer, S., Kızılhan, H., M. Karaköse, M.: Sağlık Alanındaki Etkileşimlerde Duygu Analizi Yaklaşımları ve Analizi, Sentiment Analysis Approaches and Analysis in Health Interactions, 2022.

[Uyaroğlu Akdeniz, 21] Uyaroğlu Akdeniz, F. N., Cebeci, H. İ.: Belediye Hizmetlerin Değerlendirilmesinde Duygu Analizi Yaklaşımı: Sakarya İli Örneği. Journal of Intelligent Systems: Theory and Applications, 4(2), 2021, 127-135.

[Vural, 13] Vural, A. G., Cambazoglu, B. B., Senkul, P., Tokgoz, Z. O.: A Framework for Sentiment Analysis in Turkish: Application to Polarity Detection of Movie Reviews in Turkish. In Computer and Information Sciences III: 27th International Symposium on Computer and Information Sciences, Springer London, 2013, 437-445.

[Wankhade, 22] Wankhade, M., Rao, A. C. S., Kulkarni, C.: A Survey on Sentiment Analysis Methods, Applications, and Challenges. Artificial Intelligence Review, 55(7), 2022, 5731-5780.

[Yalman, 22] Yalman, H. İ., Tüfekci, Z.: Konuşma Tanımaya Uygulanan BiRNN, BiLSTM ve BiGRU Modellerinin Performans Değerlendirmesi. Avrupa Bilim ve Teknoloji Dergisi, (36), 2022, 121-127.