www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 13 Issue 11 November 2024, Page No. 26622-26640
ISSN: 2319-7242 DOI: 10.18535/ijecs/v13i11.4932

Sentiment Analysis Using Machine Learning Methods on News Texts

Aslı Gizem Öztürk [1], Metin Turan [1]

[1] Istanbul Commerce University

Abstract
The categorization of intricate emotional phenomena in news texts permits the assessment of the impact of such news on society. In this context, it is evident that the existing body of research on emotion analysis in Turkish texts is insufficient. The objective of this study is to identify the emotions of "fear," "happiness," "anger," "sadness," and "surprise" in Turkish news texts using machine learning methods. In consideration of the Turkish language structure, the adjectives and nouns were parsed on a specially prepared corpus and dictionary groups specific to this study were created with the use of the TF-IDF and Double Normalization methods. Subsequently, these lexicon groups were employed for the classification of sentiment in news texts using a range of machine learning techniques, including Multilayer Perceptron (MLP), Support Vector Machines (SVM), Random Forest (RF), Light GBM, XG Boost, GPT-2, BERT, Distil BERT, and XLM ROBERT a. The evaluation metrics employed were F1-score, accuracy, recall, and precision. In the training phase of the model, the dictionary group created with Double Normalization, the GPT-2 model, the Grid Search optimization method, and cross-validation were employed in order to achieve the highest F1-score of 0.99 and accuracy of 0.99. It is anticipated that the methods and dataset utilized in this study will contribute to the existing literature on sentiment classification in Turkish texts and provide insights that will inform future similar studies.

Keywords: Sentiment analysis, GPT-2, XLM, natural language processing

1. Introduction

In the contemporary era, the pervasive use of technology and the internet has facilitated rapid access to information, conferring significant convenience. However, this situation has also given rise to the potential for the uncontrolled dissemination of content that may have a detrimental impact on social mental health. It is well documented that news content, particularly that which pertains to violence, fear, and tragic occurrences, can evoke a range of emotions, including anxiety, anger, and insecurity, within society. In light of these considerations, it has become imperative to safeguard both individual and social psychological well-being by subjecting news content to an evaluation process that classifies emotions prior to its publication (Kazan, 2021). The application of machine learning and natural language processing (NLP) techniques enables the automatic analysis of emotional tendencies in large-scale textual data. In this study, we employ a machine learning-based sentiment analysis approach to identify basic emotions, including fear, happiness, anger, sadness, and surprise, in Turkish news texts. The emotional analysis of news content is regarded as a crucial step in the prevention of potential negative outcomes. (Birjali et al., 2020) defined sentiment analysis as a task within the field of natural language processing (NLP), which aims to extract emotions and thoughts from texts. In its nascent stages, this technique operated by examining the emotional valence (positive, negative, or neutral) conveyed by words and phrases. Sentiment analysis is particularly advantageous in the context of large datasets, such as those comprising customer reviews or social media posts. This is because it enables the extraction of general trends within texts or the identification of public opinion on a given topic. Sentiment analysis employs statistical NLP techniques that examine the relationships between words in order to identify emotional nuances in textual data. In their review, (Mäntylä et al., 2018) noted that the origins of sentiment analysis can be traced back to 1937, when a scientific journal on public opinion was established. During this period, research and methods for detecting the emotional aspects of texts began to be developed. However, there has been a significant increase and advancement in sentiment analysis and opinion mining in recent years. The number of papers published in this field has increased exponentially, particularly after 2004. Automatic sentiment detection in texts has gained popularity since 2010 (Acheampong et al., 2020). Consequently, sentiment analysis has become a rapidly growing research area in recent years (Arazy & Woo, 2007).

A review of the studies conducted in the Turkish language reveals that emotion analysis is predominantly categorized into three distinct categories: positive, negative, and neutral. These methods are insufficient for accurately reflecting the emotional state in text analysis. The present study demonstrated that the emotions expressed in these texts can be classified into distinct categories that are analogous to the classification of facial expressions. Moreover, our findings demonstrated that the discriminability between emotion transitions can be accurately classified by integrating insights from prior studies with a state-of-the-art approach. In the feature extraction approach, a distinctive corpus was constructed using the method of adjective-noun decomposition, which differs from other methods used in this field. In the context of sentence or text analysis, emotional states are defined as adjective elements qualifying nouns. Consequently, the parsing of sentences into nouns and adjectives for the purposes of feature extraction yielded a smaller number of outputs of greater value. In this context, feature extraction, double normalization, term frequency-inverse document frequency (TF-IDF), and machine learning methods were employed to analyze and classify news texts. Double normalization is not a standard or widely utilized technique. Theoretically, by reducing the variability in the number of features, the transformer-based language model yielded more accurate results compared to other machine learning models.This study is structured as follows: Section 2 presents an overview of previous studies on sentiment analysis in texts; Section 3 outlines the materials and methods used in our study, including details on the data set; Section 4 discusses the machine learning models employed in text analysis; Section 5 presents the experimental results and evaluation of the models used in the previous section; and Section 6 offers a discussion and conclusions.

2. Related Works

Sentiment analysis with natural language processing (NLP) is the field of automatically identifying and classifying emotions expressed in texts. Work in this area has attracted significant attention in recent years and has been used in a multitude of applications. Initially, sentiment analysis studies focused on understanding the presence of basic emotions internally. In other words, the presence of emotion was classified as positive, negative, or neutral. The following are examples of this classification: (Türkmenoglu & Tantug, 2014) conducted a comparative analysis of dictionary-based and machine learning-based sentiment analysis using a dataset comprising Twitter posts and movie reviews. The emotions to be classified are divided into four categories: positive, negative, neutral, and indifferent. In his machine learning-based sentiment analysis study, he employed TF-IDF feature extraction. The Support Vector Machines (SVM) algorithm demonstrated the best performance on the Twitter dataset, achieving 85% accuracy. His findings indicated that machine learning-based sentiment analysis outperformed dictionary-based sentiment analysis.

(Akba et al., 2014) conducted a sentiment analysis study on texts pertaining to film criticism. The texts were classified as positive, negative, or neutral. The researchers employed chi-squared and Information Gain feature extraction. Feature selection was conducted in two stages, with 375 and 1,000 features, respectively.

They utilized machine learning and dictionary-based methods. The SVM classifier yielded 375 feature extractions and an F1-score of 0.63, which outperformed the Naïve Bayes (NB) classifier. In a sentiment analysis study conducted by (Velioğlu et al., 2018), Turkish tweets were classified as positive, negative, or neutral based on the presence of emojis. The authors employed a variety of machine learning algorithms, including Naïve Bayes (NB), Logistic Regression (LR), SVM, and Decision Tree. They utilized two distinct approaches, one based on bag of words and the other on fastText. Their findings revealed that both approaches yielded comparable results. When LR was applied with TF-IDF, the success rate reached 78%. In their study, (Shehu & Tokat, 2020) preprocessed 13 thousand Turkish tweets for sentiment analysis and classified them as positive, negative, or neutral using SVM and RF classification algorithms. While the RF algorithm proved sufficient for positive classification, the SVM algorithm demonstrated greater success in negative and neutral classification. The hybrid approach was observed to improve sentiment analysis performance compared to a single method. In their 2020 study, (Onan, 2017) employed a range of machine learning algorithms, including NB, SVM, and LR, on data labeled as positive and negative to analyze the sentiment of messages written in Turkish on Twitter. The highest performance value was observed when 1-gram and 2-gram feature extraction was used with the NB algorithm, yielding a value of 0.77.

In a sentiment analysis study conducted by (Demirci et al., 2019) Turkish tweets were preprocessed and labeled as positive or negative. Four machine learning models were employed: logistic regression (LR), support vector machines (SVM), random forest (RF), and deep learning (DL). The RF model, trained with term frequency–inverse document frequency (TF-IDF) features, achieved an accuracy of 81.86%. Subsequently, the classification of emotions was employed as a means of addressing more intricate issues. In this regard, an examination of prior studies is undertaken. (Asghar et al., 2019) employed a range of machine learning techniques, including random forest (RF), support vector machine (SVM), logistic regression (LR), extreme gradient boosting (XGBoost), stochastic gradient descent (SGD) classifier, naive Bayes (NB), and k-nearest neighbor (k-NN), for the purpose of emotion detection and classification in English texts. The 5,477-entry International Study on Emotion Antecedents and Reactions (ISEAR) dataset was utilized for this purpose. The study concentrated on the emotional states of fear, guilt, joy, shame, and sadness. The highest level of accuracy was achieved with the LR method, at 66.58%. The SVM method yielded an accuracy of 64.66%, while the LR method achieved 66.58%. The k-NN method yielded an accuracy of 57.81%, the NB method produced 63.6%, the RF method resulted in 64.02%, the XGBoost method achieved 58.54%, and the SGD Classifier method attained 65.57%.

(Nasir et al., 2020) conducted a study on the ISEAR dataset with the objective of identifying the relative strengths and weaknesses of different machine learning classifiers. The researchers examined the emotional states associated with anger, shame, disgust, guilt, and sadness. The classification was conducted using the support vector machine (SVM), k-nearest neighbor (k-NN), and decision tree (DT) algorithms. The performance of the algorithms was evaluated using metrics such as precision, recall, and F1-score. The NB algorithm demonstrated the highest performance, with an accuracy of 64.08%. The performance of the optimal model, which included TF-IDF and count vectorizer, as well as a confusion matrix, precision-recall ratio, and ROC (receiver operating characteristic) score, was also analyzed in detail. In their study, (Hasan et al., 2019) employed a system called Emotex to construct models for the classification of emotion. The experimental results demonstrate that these models exhibit a 90% accuracy rate in classifying sentiment in text messages. A two-stage framework, designated as Emotex Stream, is employed for the purpose of classifying live tweet streams in real time for the purpose of tracking the sentiment expressed. In the initial stage, tweets that contain explicit emotion are separated from those that do not. In the subsequent stage, a fine-grained sentiment classification is performed on the aforementioned tweets that contain explicit sentiment, utilizing Emotex. SVM, DT, and NB classification algorithms are employed for the purpose of classifying the emotions of happiness and activity, sadness and activity, happiness and inactivity, and sadness and inactivity, as previously described. (Güven et al., 2018) investigated an n-stage Latent Dirichlet Discrimination (LDA) model for the emotional classification of Turkish tweets. The analysis revealed that tweets exhibited five distinct emotional states: anger, fear, happiness, sadness, and surprise. In the system, the classical LDA method is compared with the LDA developed with N (2 and 3) stages. The results demonstrated a notable increase in success compared to the classical LDA method. The most significant factor contributing to this improvement was identified as the reduction in the number of words utilized in all documents, resulting from the elimination of words with low weight.

(Balli et al., 2022) conducted a sentiment analysis study employing a variety of machine learning algorithms on Turkish datasets gathered from Twitter and labeled as positive, negative, and neutral. The data underwent preprocessing stages through the Zemberek library and the SnowBall library. The LR, SVM, NB, RF, and SGD algorithms were utilized for sentiment analysis, resulting in classification performance with up to 87% accuracy.In a recent study, (Boynukalin & Karagoz, 2013) analyzed the sentiment of 1,161 Turkish children's tales and 4,265 sentences translated from the ISEAR dataset. They employed machine learning methods to examine the data labeled as anger, fear, joy, and sadness. The most effective approach was the NB algorithm, which achieved an 80% success rate. In a study employing machine learning methods, (Demirci, 2015) analyzed a dataset comprising 6,000 instances of six basic emotions: fear, anger, surprise, sadness, joy, and disgust, with 1,000 instances of each. The study utilized three classification algorithms, namely, Naïve Bayes, SVM, and k-NN. The success rate achieved was approximately 70%.

In their study, (Tocoglu & Alpkocak, 2018) created a new dataset, TREMO, comprising the emotion categories of fear, happiness, anger, sadness, surprise, and disgust, for use in Turkish sentiment analysis. The dataset was subjected to a series of preprocessing steps and employed in classification algorithms, including SVM, RF, and NB. The SVM algorithm demonstrated the most optimal performance, achieving an accuracy of 86.29%. (Tocoglu & Alpkocak, 2019) put forth a proposal for a Turkish dictionary encompassing the emotion categories of fear, happiness, anger, sadness, surprise, and disgust. The TREMO dataset served as the foundation for the dictionary, with Turklemma and Zemberek employed to create the terms. It was observed that Turklemma exhibited superior performance. The study demonstrated that the proposed dictionary effectively produces comparable results for sentiment analysis in Turkish texts. (Tocoglu et al., 2019) conducted a sentiment analysis of Turkish tweets using neural networks. The dataset consisted of approximately 195,000 data points representing the emotion categories of fear, happiness, anger, sadness, surprise, and disgust. Three distinct deep learning architectures were employed: artificial neural networks (ANN), convolutional neural networks (CNN), and long short-term memory networks (LSTM). The CNN architecture demonstrated the highest accuracy, reaching 74%. This study revealed that deep neural networks exhibit superior scalability with larger data sets compared to traditional machine learning algorithms.

(López-Chau et al., 2020) employed sentiment analysis and supervised learning based on six emotional models to examine datasets of trending topics on Twitter that emerged from Mexican citizens' interactions during the September 19, 2017 earthquake. Three classifiers were created to identify the emotions expressed in tweets pertaining to a specific topic. The classifiers that demonstrated the highest accuracy in predicting emotions were NB and SVM. The most frequently predicted emotions were happiness, anger, and sadness, and 6.5% of the predicted tweets were neutral. The range of emotions was extended from three (negative, neutral, positive) to six to provide a more comprehensive understanding of how users interact with social media platforms. In the study conducted by (Nasir et al., 2020) text-based sources were identified as a prevalent avenue for individuals to disseminate their opinions and emotional responses regarding a product or service through various online platforms, including social media and e-commerce sites. It is relatively simple for people to misinterpret emotions, particularly those derived from text. The primary objective of this study is to develop a text-based sentiment recognition and prediction system. In the development of sentiment analysis, various market challenges are encountered, with accuracy being the primary issue. Therefore, four supervised machine learning classification algorithms, including Multinomial NB, SVM, DT, and k-NN, are investigated. The primary objective of the research conducted by (Yohanes et al., 2023) is to examine and analyze the fundamental emotional categories that are commonly exhibited by humans, including sadness, joy, anger, fear, and disgust. The implementation of emotion detection in business-oriented sectors presents a multitude of potential advantages, including the provision of personalized services and the development of targeted mental health treatments in medical contexts. In order to construct an emotion detection model, it is essential to conduct an analysis of the most effective model currently available. This paper examines the relative performance of three recurrent neural network (RNN) models: long short-term memory (LSTM), bidirectional long short-term memory (Bi-LSTM), and gated recurrent units (GRU). The findings of the study indicate that the most straightforward RNN model, GRU, attains the highest score across four evaluation criteria: accuracy, recall, precision, and F1-score. The GRU model demonstrated a 60.26% accuracy rate, while the Bi-LSTM and LSTM models exhibited 59.3% and 57.65%, respectively. In a recent publication, (Sam Abraham et al., 2022) put forth a novel approach to emotion detection, proposing a single/multi-class or multi-label classification task for this purpose. Prior research has not examined or quantified the contribution of sentiment words or named entities to the task of detecting readers' emotions. In this study, a bidirectional long short-term memory (Bi-LSTM) network and an attention model are employed to examine the interpretable characteristics and behavior of the model for readers' emotion detection through multi-target regression settings on short-text news documents. Moreover, the study utilized six fundamental emotions (happiness, sadness, anger, fear, disgust, and surprise) to represent readers' emotional states. These are the most commonly discussed basic emotions in discrete emotion models by theorists, and most social media platforms allow users to express their emotional reactions to news or posts in different ways.

3. Material and Method

3.1 Data Set

The data set was obtained from the Hürriyet newspaper website via web crawling, a process by which digital content is extracted from a website. The data set consists of 5,000 news items that were prepared by the editors. In order to classify the news items in the dataset, the software produced by ParallelDots, a company focusing on artificial intelligence research, namely Komprehend, has been found to be reliable by more than twelve thousand developers (Komprehend, 2020). The software is utilized by over twelve thousand developers and employed by major global companies. The Komprehend data tagging tool is employed for the purpose of tagging English texts. The initial step involved translating the Turkish texts into English and providing them to the data labeling tool. The initial emotion detection and classification was conducted using the aforementioned application. Upon validation with the ChatGPT4 model, the results were found to be inconsistent. The labels were then manually checked one by one by a psychologist, and the misclassifications were corrected to a significant extent. Following this correction, the success rate F1-score increased from 0.60 to over 0.90. The news items in the dataset were then categorized into five distinct classes, with the number of items and the classes themselves presented in Table 1.

Table 1. Number of emotions classified

Classification Number	Emotion	Number of Data
1	Fear	3355
2	Surprise	2373
3	Anger	1697
4	Happiness	837
5	Sadness	718

The number of data points could not be balanced due to the presence of emotion-based news. Accordingly, different models were run in order to accommodate the imbalance in this numerical distribution, as previously mentioned in the Models section. Our dataset was prepared in Excel format and imported into Python using the Pandas library. The class-based sample of the special corpus created for our study is shown in Table 2.

Table 2. Data set sample

Emotion States	News Text
Fear	Dev kasırga hızla sahile yaklaşıyor, yetkililer halkı acilen tahliye olmaya çağırıyor. Son tahliye saatlerine girilirken endişe ve kaos büyüyor.
Surprise	Beklenmedik bir patlama, şehrin merkezinde büyük paniğe yol açtı. Yetkililer olay yerini güvenlik altına alırken, caddenin tamamen kapatıldığı bildirildi.
Anger	Şirket içinde yaşanan büyük tartışmanın ardından bir çalışan ani bir şekilde işten çıkarıldı. Kararın haksız olduğunu düşünen çalışan, hakkını aramak için mahkemeye başvurdu.
Happiness	Milli takım tarihi bir başarıya imza attı! Zorlu geçen turnuvanın ardından finalde rakibini yenerek şampiyon oldu. Ülke genelinde büyük sevinç dalgası yaşanıyor.
Sadness	Sevilen sanatçının ani vefat haberi tüm ülkeyi yasa boğdu. On binlerce hayranı sosyal medya üzerinden taziye mesajları paylaşarak, derin üzüntülerini dile getirdi.

3.2 Data Preprocessing

Data pre-processing is the process of preparing raw data for analysis and modeling. It can be considered analogous to the process by which a sculptor transforms marble into a sculpture. This stage encompasses techniques such as enhancing the data's interpretability, improving data quality, and reducing the number of dimensions, which enhance the efficiency and precision of modeling. The applied processes are illustrated in Figure 1.

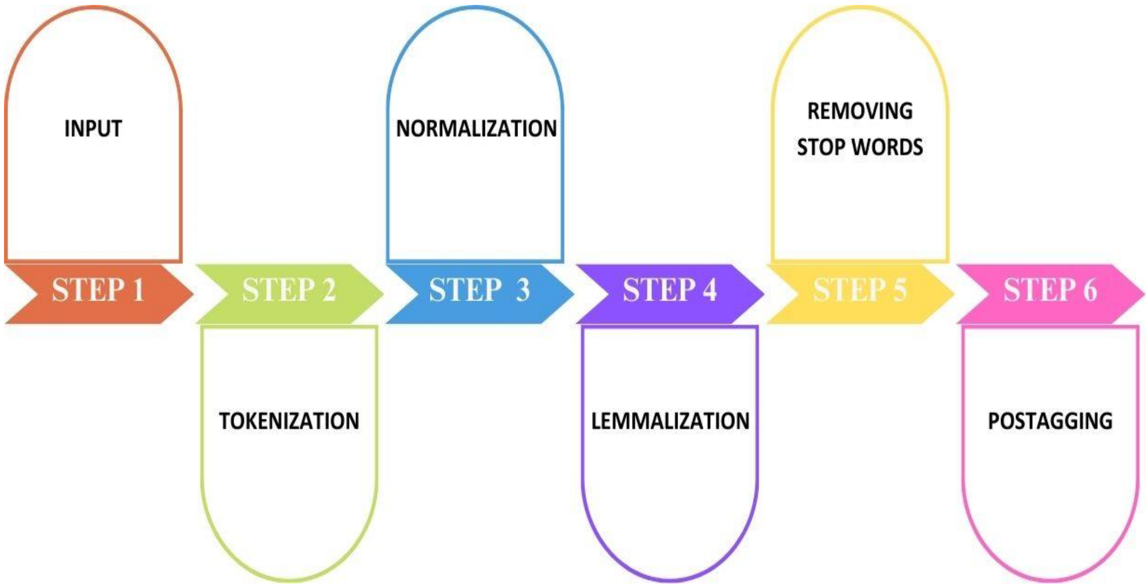

Figure 1. Data preprocessing stages

3.3 Methods

Following the preprocessing of the documents in the corpus, a dataset of adjectives and nouns (the words that best represent emotional expressions) was constructed. The words were then converted to lowercase, stemmed, and purged of stop words. The data was subjected to training through the use of K-fold cross-validation methods, with three, five, or ten folds, respectively. In the preliminary evaluation, the results that demonstrated the greatest consistency were subjected to further analysis using K-fold 10. The term "TD-IDF" refers to a statistical method that is used to indicate the relative importance of a given term within a document. In the context of text processing, double normalization refers to a two-stage normalization procedure. This entails dividing the raw frequency of a given term by the maximum raw term frequency in order to normalize the ratio of that term to other terms, regardless of the document length. The values of TF, TF-IDF, normalization, and double normalization for the words in the corpus were calculated using the equations presented in (1-6).

$$Ham\ Frekans = f_{t,d} \tag{1}$$

$$\text{Normalizasyon} = \log\left(1 + f_{t,d}\right) \tag{2}$$

$$Double\ Normalizasyon\ 0{,}5 = 0{,}5 + 0{,}5\frac{f_{t,d}}{maxf_{t,d}} \tag{3}$$

$$Terim\ Sıklığı\ TF(t,d) = \frac{t\ nin\ dokumanda\ görünme\ sayısı}{dokümandaki\ toplam\ terim\ sayısı} \tag{4}$$

$$Ters\ Belge\ Sıklığı\ IDF(t) = \log\left(\frac{N}{1+df}\right. \tag{5}$$

$$Terim\ ve\ Ters\ Doküman\ Frekansı\ TF-IDF(t,d) = TF(t,d) * IDF(t) \tag{6}$$

The 50 most salient words, which collectively represent the entire corpus, were identified. The 50 words were composed of 10 unique words from each class with high TF-IDF/double normalization values, as determined by the analysis. Nevertheless, it was determined that the 50 words were insufficient for differentiating between emotional states. Subsequently, the impact of augmenting the number of corpus representation words to 250, 500, and 750 was investigated. In order to utilize the texts within the dataset as a vector in machine learning models (MLP, SVM, RF, LightGBM, XGBoost, GPT-2, BERT, DistilBERT, XLM RoBERTa), the digitalization of the label data was performed with the Label Encoder method. The Label Encoder was utilized to transform the categorical data representing the five basic emotions (fear, happiness, anger, surprise, and sadness) into numeric values (0, 1, 2, 3, and 4). This was done to facilitate the use of these emotions as inputs to machine learning algorithms.

The TF-IDF and double normalization methods were used to transform the models into vector matrices, which could then be fed into the models. The general structure of the model designed for this purpose is shown in Figure 2.

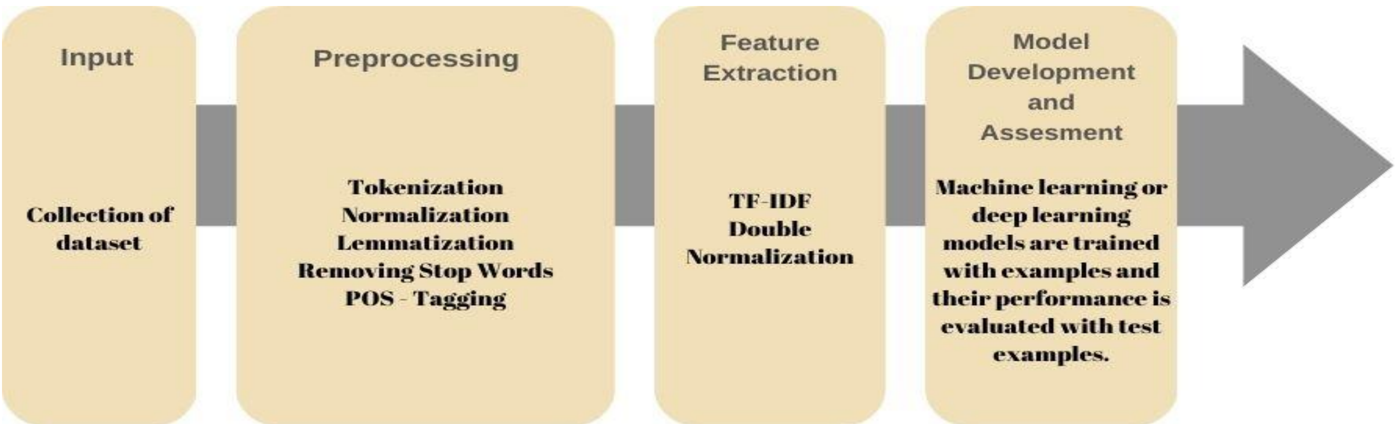

Figure 2. Sentiment analysis model structure

3.4 Hyperparameter Selection

The selection of hyperparameters is a crucial step in optimizing the performance of a machine learning model. Hyperparameters are parameters that are not adjusted during the learning process of the model; rather, they are determined prior to training. If inappropriate hyperparameters are chosen, the model may exhibit issues such as overfitting or underfitting. In this study, we selected the Grid Search method, which employs a comprehensive search strategy through systematic evaluation of all potential combinations of hyper-parameters, thereby ensuring effective results within limited search spaces. The Grid Search method was applied to all models in conjunction with K-fold, thus facilitating optimal parameter selection.

4. Models

In the study, two distinct numerical distributions were employed for the purpose of classification, as illustrated in Table 3. The K-fold (3-5-10) cross-validation method was employed in the dataset utilized for model construction. Given the nature of the problem, which is prone to overlearning (multi-class and overlapping emotions), it is more meaningful to evaluate the model by testing it in a manner that is as close to reality as possible. The aim was to ensure that the evaluation was free from misconceptions by ensuring that the test results were realistic and that the test sample was sufficiently separated. Consequently, the results obtained reflect a model that can be adapted to real-life scenarios. The number of data points utilized for each emotion class in the models is presented in Table 3.

Table 3. Number of data used according to models

Emotions	Model-1	Model-2
Fear	1000	700
Surprise	1000	700
Anger	1000	700
Happiness	1000	700
Sadness	1000	700
Total	**5000**	**3500**

In the initial phase of the sentiment analysis study, Model-1 was employed to train the model on a corpus comprising 5,000 equally classified sentiments, with an equal number of instances of each sentiment. It was observed that data with a balanced class distribution affected the model's success rate. In Model-2, the corpus was again reduced to 3,500 by organizing it to include equal distribution for all emotions, and the training model was performed. For the corpus created in each model, TF-IDF and double normalization calculations were performed again, and dictionaries were created. Feature extraction was performed by separating adjectives and nouns from the texts.

4.1 Multilayer Perceptron Neural Network Classifier (MLP)

Artificial neural networks (ANNs) are powerful artificial intelligence (AI) tools that are widely used in various fields, including diagnosis, classification, prediction, control, data association, data filtering, interpretation, and others. The selection of the optimal neural network architecture for addressing a specific problem should be based on a comprehensive evaluation of the distinctive attributes of the network and the problem at hand. One potential solution is a multilayer neural network (MLP), a variant of the artificial neural network. An MLP is a feed-forward artificial neural network model comprising input, hidden, and output layers. The layer structure of the MLP model utilized in this study is presented in Figure 3.

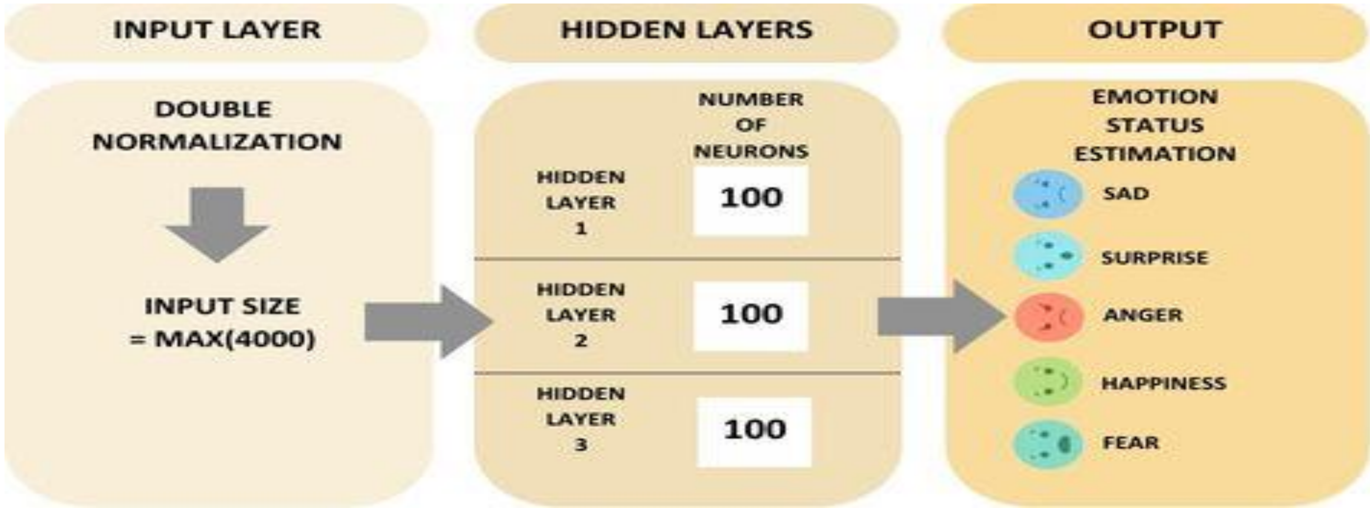

Figure 3. Illustrates the structure of the MLP classifier artificial neural network model.

In the input layer, words obtained through double normalization feature extraction are employed as input. Each neuron in this layer is assigned a weight, which represents the relative importance of a word or term. Upon analysis of the hidden layer, it became evident that there are three distinct hidden layers, designated as H1, H2, and H3. Each of these layers contains 100 neurons. Each neuron is configured to receive signals from all neurons in the preceding layer and subsequently pass them through an activation function. The model employs the ReLU (Rectified Linear Unit) activation function. In the output layer, the softmax activation function is employed to predict the classification of each text. During training, the model updates its weights using the training data with the objective of optimizing its performance on the test data. The loss function "categorical cross-entropy" is utilized to train the training model.

4.2 Bidirectional Encoder Representations with Transformers (BERT)

In 2018, Google AL developed BERT (Bidirectional Encoder Representations from Transformers), a highly successful deep learning model for language modeling and natural language processing (NLP). In contrast to conventional language models, BERT considers bidirectional dependencies between words by processing both left-to-right and right-to-left inputs. BERT is employed in a multitude of NLP applications, including search engines, social media analysis, chatbots, and text summarization (Devlin, 2018). Upon examination of the outcomes yielded by BERT, it was observed that the results were nearly identical to those obtained by the MLP model.

4.3 Support Vector Machines (SVM)

SVM is a robust and versatile algorithm that offers solutions to classification and regression problems within the supervised learning paradigm. The fundamental premise of SVM is to discern a hyperplane that optimally differentiates the data. This hyperplane is delineated by specific data points, designated as support vectors, which provide the greatest margin between data points. The attributes of SVMs, including high generalization ability, low variance, and a minimal number of tunable parameters, have rendered them a preferred choice in numerous domains. SVMs are extensively utilized in domains such as image recognition, text classification, and outlier detection (Cortes, 1995).

4.4 XGBoost

XGBoost (Extreme Gradient Boosting) represents a variant of the gradient boosting algorithm, optimized for enhanced speed and efficiency. Gradient boosting is predicated on the principle of rectifying errors through the sequential training of a set of weak prediction models, typically decision trees. XGBoost represents an enhanced iteration of this process, offering accelerated speeds and enhanced performance through the application of diverse optimization techniques. XGBoost is a frequently preferred model in data science competitions, such as those held by Kaggle. This is due to its balanced performance in terms of both accuracy and computational cost. In the context of academic research, XGBoost is frequently employed for classification and regression analysis. The model is notable for its rapid and effective results on large datasets in areas such as finance, health, bioinformatics, and marketing analytics (Murphy, 2006). In the context of classification problems, the log loss function, as illustrated in Equation 7, is utilized.

$$L(\theta) = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) \quad (7)$$

- L(θ): The overall loss function of the model.
- $\ell(y_i, \hat{y}_i)$: The loss between the true value (y_i) and the model predicted value ($\hat{y}_i$) for the i-th observation

XGBoost differs from the gradient boosting algorithm in that it employs regularization terms to regulate the complexity of the model. These terms serve the purpose of preventing the model from overfitting. The regularization term can be identified in Equation 8.

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda j \sum_{j=1}^{T} w_j^2 \quad (8)$$

- T : Number of leaves on the tree.
- w_j: Prediction weights at each leaf node.
- γ : Parameter controlling the cost of the number of leaves.
- λ : The regulation coefficient penalizes large w_j values.

4.5 Random Forest (RF)

The RF classifier is a machine learning algorithm that has been constructed by combining multiple decision trees. The construction of each decision tree involves the training process being carried out on a randomly selected subset of features. As a consequence of this approach, the RF classifier generates a model that is more resilient and exhibits a reduced propensity to err in comparison to a single decision tree (Breiman, 2001).

4.6 OpenAI GPT-2

The OpenAI GPT-2 model was initially proposed by Radford A. in the seminal text, "Language Models are Unsupervised Multitask Learners" (Radford et al., 2019). It is a one-way transformer that has been pre-trained using language modeling on a substantial corpus of text data, comprising 40 gigabytes. GPT-2 (Generative Pretrained Transformer 2) is a self-supervised language model comprising 1.5 billion parameters and constructed on the Transformer architectural framework. The Transformer architecture is capable of understanding the context of language through the use of attention mechanisms, which facilitate the interpretation of linguistic data in a contextualized manner. In this context, GPT-2 has been trained on a substantial text dataset. Notwithstanding the absence of fine-tuning for a specific task, it evinces proficiency in a plethora of natural language processing tasks. It is noteworthy that the model is capable of attaining high performance with minimal data. This is particularly advantageous when limited data is available in specific domains. Additionally, GPT-2 is highly proficient in ensuring the consistency of texts in terms of grammar, meaning, and flow, thereby enabling its use in academic writing processes. GPT-2 is an autoregressive language model that predicts each word based on previous words. The mathematical representation of this process is expressed in Equation 9 for a sequence of words x1, x2, ..., xT given in the language model:

$$P(x_1, x_2, \ldots, x_T) = \prod_{t=1}^{T} P(x_t \mid x_1, x_2, \ldots, x_{t-1}) \quad (9)$$

- $P(x_1, x_2, ..., x_T)$: The probability of a given word sequence.
- $P(x_t \mid x_1, x_2, ..., x_{t-1})$: The probability that the model predicts the t-th word in the context of previous words.

The self-attention mechanism, which constitutes the fundamental element of the transformer structure, is illustrated in Equation 10:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (10)$$

- Q(Query), K(Key), ve V(Value): They are vectors that are inputs to the model.
- d_k: Vector dimension (used for normalization purposes).
- Softmax: It produces a probability distribution by calculating the probability of each item.

The self-attention mechanism enables the model to discern the contextual nuances of each word and identify which words are of greater significance. Figure 4 illustrates the GPT-2 architectural framework of our classification model.

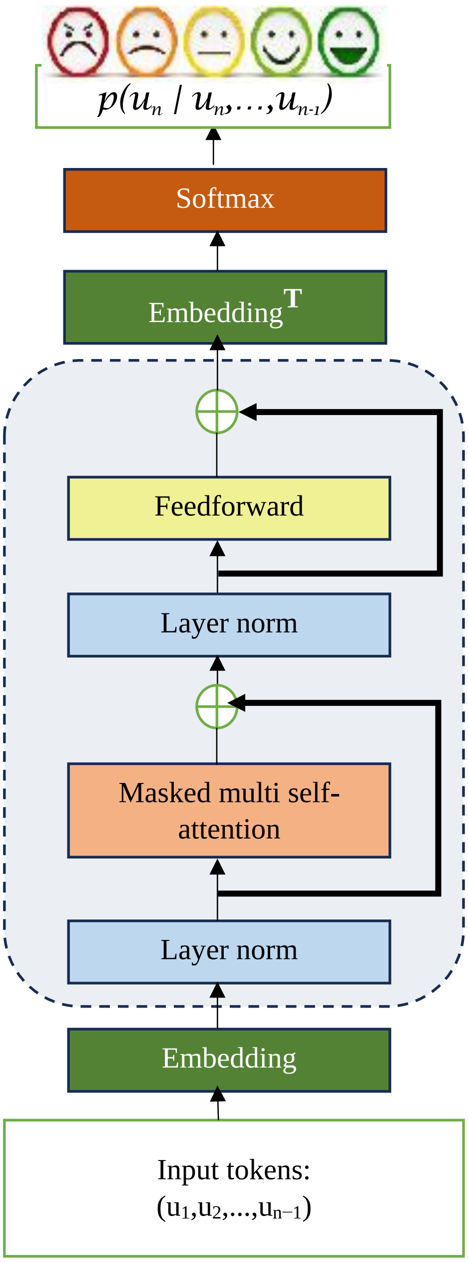

Figure 4. Sentiment analysis GPT-2 architecture

4.7 LightGBM

Light GBM (Light Gradient Boosting Machine) is an open-source machine learning library developed by Microsoft. Light GBM represents an optimized version of the Gradient Boosting Decision Tree (GBDT) algorithm, offering high performance, particularly when working with large datasets and problems characterized by high-dimensional features (GuolinKe et al., 2017). This model offers notable performance advantages over traditional variants of GBDT through the use of a histogram-based algorithm, which facilitates more rapid training processes and reduced memory usage. LightGBM is distinguished by its parallel processing capabilities, support for handling missing data, automatic categorization of features, and its effectiveness across a range of tasks, including classification, regression, ranking, and time series analysis. Light GBM employs an incremental approach to tree-based modeling, as illustrated in Equation 11, whereby at each stage, a new tree is constructed to refine the previous model and reduce error.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{11}$$

- $\boldsymbol{F_m(x)}$: m-the model in the first stage. This refers to the sum of existing trees.
- $\boldsymbol{F_{m-1}(x)}$: m-1-the model in the first step, i.e. the model learned in the previous step.
- $\boldsymbol{h_m(x)}$: decision tree added in the m-th stage.
- $\boldsymbol{\gamma_m}$: coefficient adjusting the contribution of tree $\boldsymbol{h_m(x)}$.
- $\boldsymbol{x}$: Input feature vector.

The objective at each stage of the process is to reduce the residual error by learning a new decision tree. This is achieved by utilising a negative gradient, which is the derivative of the error function. This is the reason behind the model being referred to as "gradient boosting".

5. Experimental Results

The hyperparameters were optimized using the Lattice Search method, and the optimal hyperparameter values for the GPT-2 model are presented in Table 4.

Table 4. Optimum parameter selection with Lattice Search method

GPT-2 Model Hyper Parameters	Optimum Value
Logging steps	10
Per device train batch size	8
Per device eval batch size	8
Num train epochs	3
Warmup steps	500
Weight decay	0,01

In this study on sentiment analysis in Turkish news texts, the highest score was achieved with a corpus of 5,000 in the OPENAI GPT-2 Model 1. Figure 5 illustrates the confusion matrix and sentiment distribution of the optimal result, demonstrating that all emotions were accurately classified with high accuracy.

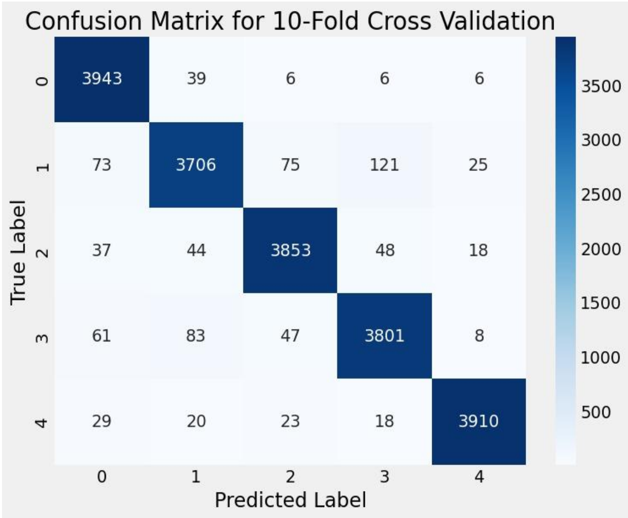

Figure 5. GPT-2 confusion matrix for Model-1

In light of the methodology, feature extraction, and corpus numbers utilized in this study, the scores derived from Models 1 and 2 with the K-fold (3-5-10) method are presented in Table 5 on an algorithmic basis.

Table 5. Scores from models by method, number of dictionaries, corpus and cross-validations

Method Used	Number of Dictionaries	Corpus Size	K-fold	F1-Score	Accuracy	Recall	Precission
SVM	250	5000	3	0,87	0,87	0,87	0,89
MLP				0,81	0,8	0,8	0,82
RF				0,87	0,87	0,87	0,89
XBOOST				0,86	0,86	0,86	0,89
LIGHT GBM				0,84	0,86	0,86	0,86
SVM	250		5	0,87	0,87	0,87	0,89
MLP				0,82	0,81	0,81	0,87
RF				0,87	0,87	0,87	0,89
XBOOST				0,85	0,85	0,85	0,88
LIGHT GBM				0,85	0,86	0,86	0,87
SVM	250		10	0,87	0,87	0,87	0,89
MLP				0,83	0,82	0,82	0,86
RF				0,87	0,87	0,87	0,89
XBOOST				0,86	0,86	0,86	0,89
LIGHT GBM				0,87	0,87	0,87	0,89
SVM	500		3	0,92	0,92	0,92	0,93
MLP				0,84	0,83	0,83	0,85
RF				0,92	0,92	0,92	0,93
XBOOST				0,91	0,91	0,91	0,92
LIGHT GBM				0,91	0,89	0,89	0,91
SVM	500		5	0,93	0,92	0,92	0,93
MLP				0,84	0,84	0,84	0,85
RF				0,92	0,92	0,92	0,93
XBOOST				0,91	0,9	0,9	0,92
LIGHT GBM				0,9	0,89	0,91	0,9
SVM	500		10	0,92	0,92	0,92	0,93
MLP				0,85	0,85	0,85	0,86
RF				0,92	0,92	0,92	0,93
XBOOST				0,91	0,91	0,91	0,92

LIGHT GBM				0,92	0,91	0,91	0,93
SVM	750		3	0,97	0,97	0,97	0,97
MLP				0,86	0,86	0,86	0,86
RF				0,97	0,97	0,96	0,97
XBOOST				0,94	0,94	0,94	0,95
LIGHT GBM				0,97	0,96	0,96	0,97
SVM	750		5	0,97	0,97	0,97	0,97
MLP				0,88	0,89	0,88	0,88
RF				0,97	0,97	0,96	0,97
XBOOST				0,94	0,94	0,94	0,95
LIGHT GBM				0,97	0,98	0,96	0,97
SVM	750		10	0,97	0,97	0,97	0,97
MLP				0,9	0,9	0,9	0,9
RF				0,97	0,97	0,97	0,97
XBOOST				0,94	0,94	0,94	0,95
LIGHT GBM				0,99	0,97	0,97	0,98
GPT-2	Auto Tokenizer		3	0,96	0,96	0,96	0,96
XLM ROBERTa				0,97	0,97	0,97	0,97
DistilBERT				0,94	0,94	0,94	0,94
BERT				0,97	0,97	0,97	0,97
GPT-2	Auto Tokenizer		5	0,98	0,98	0,98	0,98
XLM ROBERTa				0,98	0,98	0,98	0,98
DistilBERT				0,94	0,94	0,94	0,94
BERT				0,98	0,98	0,98	0,98
GPT-2	Auto Tokenizer		10	**0,99**	**0,99**	**0,99**	**0,99**
XLM ROBERTa				0,97	0,97	0,97	0,97
DistilBERT				0,94	0,94	0,94	0,94
BERT				0,92	0,92	0,92	0,92
SVM	250	3500	3	0,85	0,85	0,85	0,86
MLP				0,81	0,8	0,8	0,82
RF				0,91	0,91	0,91	0,92
XBOOST				0,88	0,88	0,88	0,9
LIGHT GBM				0,9	0,9	0,9	0,91
SVM	250		5	0,85	0,85	0,85	0,86
MLP				0,82	0,81	0,81	0,87
RF				0,9	0,9	0,9	0,91
XBOOST				0,9	0,9	0,9	0,91
LIGHT GBM				0,9	0,89	0,9	0,91
SVM	250		10	0,85	0,85	0,85	0,86
MLP				0,83	0,82	0,82	0,86
RF				0,91	0,9	0,9	0,92
XBOOST				0,88	0,88	0,88	0,9
LIGHT GBM				0,91	0,91	0,91	0,92
SVM	500		3	0,89	0,89	0,89	0,9
MLP				0,84	0,83	0,83	0,85
RF				0,95	0,94	0,94	0,95
XBOOST				0,92	0,92	0,92	0,92
LIGHT GBM				0,91	0,91	0,91	0,92

SVM	500		5	0,89	0,89	0,89	0,9
MLP				0,84	0,84	0,84	0,85
RF				0,95	0,95	0,95	0,95
XBOOST				0,93	0,93	0,93	0,94
LIGHT GBM				0,91	0,91	0,91	0,91
SVM	500		10	0,89	0,89	0,89	0,9
MLP				0,85	0,85	0,85	0,86
RF				0,95	0,95	0,95	0,95
XBOOST				0,93	0,93	0,93	0,93
LIGHT GBM				0,93	0,93	0,93	0,94
SVM	750		3	0,96	0,96	0,96	0,96
MLP				0,87	0,87	0,87	0,88
RF				0,97	0,97	0,97	0,97
XBOOST				0,96	0,96	0,96	0,96
LIGHT GBM				0,96	0,96	0,96	0,96
SVM	750		5	0,95	0,95	0,95	0,96
MLP				0,86	0,87	0,87	0,87
RF				0,98	0,98	0,98	0,98
XBOOST				0,96	0,96	0,96	0,96
LIGHT GBM				0,95	0,95	0,95	0,96
SVM	750		10	0,93	0,93	0,93	0,93
MLP				0,82	0,82	0,82	0,83
RF				0,97	0,97	0,97	0,97
XBOOST				0,96	0,96	0,96	0,96
LIGHT GBM				0,96	0,96	0,96	0,96
GPT-2	Auto Tokenizer		3	0,9	0,91	0,91	0,91
XLM ROBERTa				0,94	0,94	0,94	0,94
DistilBERT				0,97	0,97	0,97	0,97
BERT				0,97	0,97	0,97	0,97
GPT-2	Auto Tokenizer		5	0,98	0,99	0,99	0,98
XLM ROBERTa				0,98	0,98	0,98	0,98
DistilBERT				**0,99**	**0,99**	**0,99**	**0,99**
BERT				0,98	0,98	0,98	0,98
GPT-2	Auto Tokenizer		10	**0,99**	**0,99**	**0,99**	**0,99**
XLM ROBERTa				0,97	0,97	0,97	0,98
DistilBERT				0,98	0,98	0,97	0,98
BERT				0,98	0,98	0,98	0,98

Table 5 can be interpreted by noting that the methods with the highest F1-score, accuracy, recall, and precision values were GPT-2 Auto Tokenizer (10-fold) and DistilBERT (5-fold). Both methods demonstrated a 99% success rate in the F1-score, accuracy, recall, and precision metrics. However, SVM (750 dictionary counts and 5-fold) and RF (750 dictionary counts and 5-fold) also exhibited a notable performance, with a success rate of approximately 97% and 98%, respectively. Nevertheless, they exhibited a slight discrepancy in performance compared to GPT-2 and DistilBERT.

6. Evaluation

The evaluation and testing phase of this study employed three distinct methodologies. Training was conducted on two distinct models with 250, 500, and 750 dictionaries. The proposed model was trained with a variety of algorithms, including MLP, SVM, RF, LightGBM, XGBoost, BERT, GPT-2, XLM ROBERTa, and DistilBERT, within each method. Subsequently, the trained model was utilized for prediction with test data, and evaluation metrics of accuracy, precision, recall, and F1-score were employed. In this context, the highest performance values obtained during the prediction process on the trained model are presented in Table 6.

Table 6. Best performance values according to models

Models	Model-1	Model-2
Best Score Model	GPT-2	GPT-2
Number of Feature Extractions	50257	50257
Scores	F1-score:0,99 Accuracy :0,99	F1-score:0,99 Accuracy :0,99
Feature Extraction Method	Double Normalizasyon + Auto Tokenizer	Double Normalizasyon + Auto Tokenizer

Upon examination of Table 5, which contains the evaluation results, it is observed that the dictionary created with the Double Normalization method as the feature extraction method exhibits the highest accuracy rate across all training algorithms. In the context of machine learning algorithms, the GPT-2 model once again demonstrated the highest level of success. The two repeated models demonstrated that the selection of the number of corpora and the number of features affected the classification scores. Upon comprehensive analysis of the evaluation results, the GPT-2 model achieved the highest score with 50,257 dictionaries in Model 1. Given the limited corpus size in Model 2, it can be inferred that the GPT-2 model exhibits superior speed and performance compared to the other models. In examining the potential explanations for the superior performance of the GPT-2 Auto Tokenizer (10-fold) and DistilBERT (5-fold) methods relative to their counterparts, it becomes evident that Transformer-based models, such as GPT-2 and DistilBERT, possess the ability to more effectively comprehend the contextual nuances of language, akin to that of large language models. These models exhibit an enhanced capacity for accurate classification, which can be attributed to their enhanced understanding of the semantic structure and meaning of language.

Transformer-based models are distinguished by their capacity to learn from extensive corpora. As the size of the dictionary and the corpus increase, these models demonstrate enhanced ability to comprehend the intricate structures of the language. As illustrated in the table, GPT-2 and DistilBERT exhibited notable performance with datasets up to 750 dictionary sizes.The GPT-2 Auto Tokenizer, which achieved an F1-score of 99% with 10-fold cross-validation, is an effective technique for validating the overall performance of the model and minimizing the risk of overlearning or overfitting the model on the dataset. The high k-value of the k-fold method (10) enabled the model to be trained with different subsets of data on each fold, thereby facilitating generalization. Metrics such as precision, recall, and F1-score are all balanced, indicating that these methods have low levels of both false positive and false negative errors. This indicates that the model not only provides correct classifications but also minimizes misclassifications.

A review of the literature and datasets of previous studies on Turkish texts revealed that sentiment analysis was typically conducted using a binary classification system, categorizing text as either positive, negative, or neutral. In view of the emotions sought to be identified in these studies, the algorithms employed, and the evaluation metrics, this study will make a pioneering contribution to the field of sentiment analysis in Turkish texts, both in terms of data source and the utilisation of state-of-the-art methodologies. Through the

Lattice Search approach, it was ascertained that the most optimal hyperparameters yielded the most favourable performance scores.

7. Discussion and Conclusion

This study is based on text classification using language models of natural language processing (NLP). In the context of emotion classification, the frequency of emotion terms within a text and the sequences in which they occur are crucial factors in determining the meaning and emotion conveyed. It is evident that the selection and application of feature extraction methods are of significant consequence with respect to the success of the classification process. In particular, the creation of a reliable corpus in Turkish and the identification of a suitable, well-annotated corpus present significant challenges. A substantial corpus of accepted texts and dictionaries exists in the English language. In order to make a unique and valuable contribution to the field and to facilitate future research, we have extracted five sentiment-domain classified corpora and dictionaries from Turkish sources. The data set obtained will provide a foundation for future studies. In addition to the smart signs used for informative and warning purposes in TV and video content channels, our study suggests the potential use of new emoji signs for detecting the predominant emotional state of the broadcast. In this study, the state-of-the-art machine learning and deep learning models demonstrated particularly impressive performance. Dictionary and feature extraction methods were of pivotal importance to the success of this endeavor. In the process of extracting the dictionary, it was recognized that emotion expressions are composed of adjectives and nouns, in accordance with the linguistic structure. Consequently, only adjective-noun pairs were extracted from the corpus, resulting in the creation of a limited number of highly valuable dictionary groups. It has been demonstrated that accelerated models with a reduced number of dictionaries achieve enhanced speed and accuracy. As models become more complex and the number of parameters increases, performance improves; however, the speed of operation declines in inverse proportion. The most effective methods were the GPT-2 Auto Tokenizer and Transformer-based models, which demonstrated superior performance in natural language processing tasks. This is attributed to their capacity for deeper contextual and semantic understanding of language. These methods demonstrate superior performance compared to other models, largely due to their sophisticated grammatical and attentional mechanisms, which are derived from extensive corpus data. In this study, the language model was trained on a dataset comprising adjective-noun pairs, with the objective of focusing attention on adjective-noun phrases. It would be beneficial for future studies to employ a greater variety of Turkish datasets and feature extraction techniques, including N-gram, word bag, word2vec, and LLM models.

References

1. Kazan, H. (2021). Dijital habercilik bağlamında trajik haberlerde etik ve bireyler üzerindeki etkisi üzerine bir araştırma. International Journal of Social Inquiry, 14(2), 661-692.
2. Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-Based Systems, 226, 107134.
3. Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. Computer Science Review, 27, 16-32.
4. Acheampong, F. A., Wenyu, C., & Nunoo‐Mensah, H. (2020). Text‐based emotion detection: Advances, challenges, and opportunities. Engineering Reports, 2(7), e12189.
5. Arazy, O., & Woo, C. (2007). Enhancing information retrieval through statistical natural language processign: a study of collocation indexing. Mis Quarterly, 525-546.
6. Türkmenoglu, C., & Tantug, A. C. (2014, June). Sentiment analysis in Turkish media. In International Conference on Machine Learning (ICML).
7. Akba, F., Uçan, A., Sezer, E. A., & Sever, H. (2014, July). Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews. In 8th European Conference on Data Mining (Vol. 191, No. 2002, pp. 180-184).

8. Velioğlu, R., Yıldız, T., & Yıldırım, S. (2018, September). Sentiment analysis using learning approaches over emojis for Turkish tweets. In 2018 3rd International Conference on Computer Science and Engineering (UBMK) (pp. 303-307). IEEE.
9. Shehu, H. A., & Tokat, S. (2020). A hybrid approach for the sentiment analysis of Turkish Twitter data. In Artificial Intelligence and Applied Mathematics in Engineering Problems: Proceedings of the International Conference on Artificial Intelligence and Applied Mathematics in Engineering (ICAIAME 2019) (pp. 182-190). Springer International Publishing.
10. Onan, A. (2017). Twitter mesajlari üzerinde makine öğrenmesi yöntemlerine dayali duygu analizi. Yönetim Bilişim Sistemleri Dergisi, 3(2), 1-14.
11. Demirci, G. M., Keskin, Ş. R., & Doğan, G. (2019, December). Sentiment analysis in Turkish with deep learning. In 2019 IEEE international conference on big data (big data) (pp. 2215-2221). IEEE.
12. Asghar, M. Z., Subhan, F., Imran, M., Kundi, F. M., Shamshirband, S., Mosavi, A., ... & Varkonyi-Koczy, A. R. (2019). Performance evaluation of supervised machine learning techniques for efficient detection of emotions from online content. arXiv preprint arXiv:1908.01587.
13. Nasir, A. F. A., Nee, E. S., Choong, C. S., Ghani, A. S. A., Majeed, A. P. A., Adam, A., & Furqan, M. (2020, February). Text-based emotion prediction system using machine learning approach. In IOP Conference Series: Materials Science and Engineering (Vol. 769, No. 1, p. 012022). IOP Publishing.
14. Hasan, M., Rundensteiner, E., & Agu, E. (2019). Automatic emotion detection in text streams by analyzing twitter data. International Journal of Data Science and Analytics, 7, 35-51.
15. Güven, Z. A., Diri, B., & Çakaloğlu, T. (2018, April). Classification of TurkishTweet emotions by n-stage Latent Dirichlet Allocation. In 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT) (pp. 1-4). IEEE.
16. Balli, C., Guzel, M. S., Bostanci, E., & Mishra, A. (2022). Sentimental analysis of Twitter users from Turkish content with natural language processing. Computational Intelligence and Neuroscience, 2022(1), 2455160.
17. Boynukalin, Z., & Karagoz, P. (2013, September). Emotion analysis on Turkish texts. In Information Sciences and Systems 2013: Proceedings of the 28th International Symposium on Computer and Information Sciences (pp. 159-168). Cham: Springer International Publishing.
18. Demirci, S. (2015). Emotion analysis on Turkish tweets (Master's thesis, Middle East Technical University).
19. Tocoglu, M. A., & Alpkocak, A. (2018). TREMO: A dataset for emotion analysis in Turkish. Journal of Information Science, 44(6), 848-860.
20. Toçoğlu, M. A., & Alpkocak, A. (2019). Lexicon-based emotion analysis in Turkish. Turkish Journal of Electrical Engineering and Computer Sciences, 27(2), 1213-1227.
21. Tocoglu, M. A., Ozturkmenoglu, O., & Alpkocak, A. (2019). Emotion analysis from Turkish tweets using deep neural networks. IEEE Access, 7, 183061-183069.
22. López-Chau, A., Valle-Cruz, D., & Sandoval-Almazán, R. (2020). Sentiment analysis of Twitter data through machine learning techniques. Software engineering in the era of cloud computing, 185-209.
23. Nasir, A. F. A., Nee, E. S., Choong, C. S., Ghani, A. S. A., Majeed, A. P. A., Adam, A., & Furqan, M. (2020, February). Text-based emotion prediction system using machine learning approach. In IOP Conference Series: Materials Science and Engineering (Vol. 769, No. 1, p. 012022). IOP Publishing.
24. Yohanes, D., Putra, J. S., Filbert, K., Suryaningrum, K. M., & Saputri, H. A. (2023). Emotion Detection in Textual Data using Deep Learning. Procedia Computer Science, 227, 464-473.
25. Sam Abraham, S., VL, L., & P Gangan, M. (2022). Readers' affect: Predicting and understanding readers' emotions with deep learning. Journal of Big Data, 9(1), 1-31.
26. Komprehend, Text Analysis APIs, 2020. [Çevrimiçi]. Available: https://komprehend.io/#emotion.
27. Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
28. Cortes, C. (1995). Support-Vector Networks. Machine Learning.
29. Murphy, K. P. (2006). Naive bayes classifiers. University of British Columbia, 18(60), 1-8.
30. Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

31. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.
32. GuolinKe, Q. M., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst, 30, 52.