POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)
URL: http://dergipark.org.tr/politeknik

Comparative analysis of traditional machine learning and transformer-based deep learning models for text classification

Metin sınıflandırması için geleneksel makine öğrenimi ve dönüştürücü tabanlı derin öğrenme modellerinin karşılaştırmalı analizi

(Author(s)): *Nazif AYDIN[1], Osman Ayhan ERDEM[2], Adem TEKEREK[3]*

ORCID[1]: 0000-0001-5645-1597

ORCID[2]: 0000-0001-7761-1078

ORCID[3]: 0000-0002-0880-7955

To cite to this article: Aydın N., Erdem O. A. and Tekerek A., "Comparative Analysis of Traditional Machine Learning and Transformer-based Deep Learning Models for Text Classification", ***Journal of Polytechnic***, *(*): *, (*).

Bu makaleye şu şekilde atıfta bulunabilirsiniz: Aydın N., Erdem O. A. ve Tekerek A., "Metin Sınıflandırması için Geleneksel Makine Öğrenimi ve Dönüştürücü Tabanlı Derin Öğrenme Modellerinin Karşılaştırmalı Analizi", ***Politeknik Dergisi***, *(*): *, (*).

Erişim linki (To link to this article): http://dergipark.org.tr/politeknik/archive

DOI: 10.2339/politeknik.1469530

Comparative Analysis of Traditional Machine Learning and Transformer-based Deep Learning Models for Text Classification

Metin Sınıflandırması için Geleneksel Makine Öğrenimi ve Dönüştürücü Tabanlı Derin Öğrenme Modellerinin Karşılaştırmalı Analizi

Highlights

- *The study compares traditional machine learning techniques with Transformer-based deep learning models for text classification.*
- *Traditional methods include Decision Tree, Naive Bayes, Random Forest, and SVM.*
- *Transformer-based models evaluated are DistilBERT, BERT, GPT-2, RoBERTa, and GPT-3.*
- *Findings reveal GPT-3's significantly higher accuracy and F1 score than other models.*
- *The study underscores the promise of Transformer-based models in text classification tasks.*

Graphical Abstract

In this study, new input data is processed by deep learning/machine learning algorithms such as DistilBERT, BERT, GPT-2, RoBERTa, GPT-3, Decision Tree, Naive Bayes, Random Forest, and SVM, and predictions are made. These predictions are evaluated based on accuracy. If the accuracy is unacceptable, the algorithms are trained with training data. After training, the models process input data, make predictions again, and re-evaluate the accuracy values. This cycle continues until a successful model is obtained.

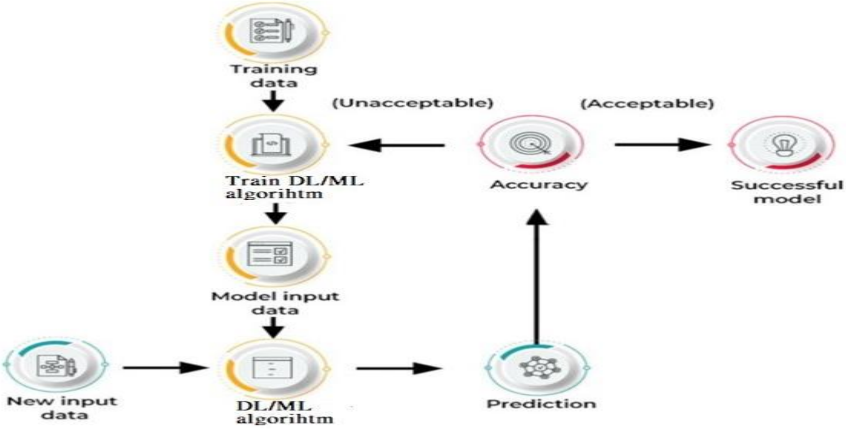

Figure. Graphical Abstract

Aim

This study aims to compare the effectiveness of traditional machine learning techniques and Transformer-based deep learning models for text classification tasks..

Design & Methodology

The study evaluates the performance of Decision Tree, Naive Bayes, Random Forest, and SVM against DistilBERT, BERT, GPT-2, RoBERTa, and GPT-3 using a dataset encompassing six educational categories.

Originality

The study contrasts traditional and innovative models in text classification, shedding light on the transformative potential of Transformer-based models.

Findings

GPT-3 demonstrates superior accuracy and F1 score compared to traditional methods and other Transformer-based models.

Conclusion

Transformer-based models, particularly GPT-3, exhibit promising efficacy in text classification, suggesting their potential as a preferred choice over traditional ML algorithms.

Declaration of Ethical Standards

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Comparative Analysis of Traditional Machine Learning and Transformer-based Deep Learning Models for Text Classification

Research Article

Nazif AYDIN[1*], Osman Ayhan ERDEM[2], Adem TEKEREK[3]

[1]Department of Computer Engineering Graduate School of Natural and Applied Sciences, Gazi University, Ankara, Türkiye
[2]Department of Computer Engineering Faculty of Technology, Gazi Universit,Ankara, Türkiye
[3]Department of Computer Engineering Faculty of Technology, Gazi Universit,Ankara, Türkiye
(Geliş/Received : 17.04.2024 ; Kabul/Accepted : 30.05.2024 ; Erken Görünüm/Early View : 05.08.2024)

ABSTRACT

In today's information age, the generation and utilization of vast amounts of textual data have become exceedingly important. Within artificial intelligence, specifically natural language processing, text classification is crucial, aiding in the organization and comprehension of this data deluge. Text classification involves categorizing text pieces and allocating them to respective classes, a process significantly advanced by machine learning and deep learning methodologies. The aim of this study is to evaluate the effectiveness of conventional machine learning algorithms, including DT, NB, RF, and SVM, alongside state-of-the-art Transformer-based models such as BERT, DistilBERT, GPT-2, GPT-3, and RoBERTa in text classification tasks. Findings indicate that while Naive Bayes achieves a 65% accuracy rate among traditional methods, GPT-3 surpasses them with a 77% higher accuracy and F1 score. These results highlight the significant promise and efficiency of Transformer-based models in text classification endeavors.

Keywords: Deep learning, machine learning, natural language processing, text classification, transformer.

Metin Sınıflandırması için Geleneksel Makine Öğrenimi ve Dönüştürücü Tabanlı Derin Öğrenme Modellerinin Karşılaştırmalı Analizi

ÖZ

Günümüz bilgi çağında, büyük miktarda metinsel verinin üretilmesi ve kullanılması son derece önemli hale gelmiştir. Yapay zeka alanında, özellikle doğal dil işleme içerisinde, metin sınıflandırma bu veri selinin düzenlenmesine ve anlaşılmasına yardımcı olan kritik bir görev olarak öne çıkar. Metin sınıflandırmanın özü, metin parçalarını kategorilere ayırarak bunları ilgili sınıflara tahsis etmektir, bu süreç, makine öğrenimi ve derin öğrenme metodlarıyla önemli ölçüde ilerletilmiştir. Bu çalışmanın amacı, Geleneksel Makine Öğrenimi teknikleri arasında Karar Ağacı, Naive Bayes, Rastgele Orman ve SVM gibi tekniklerin etkinliğini değerlendirmek, ayrıca DistilBERT, BERT, GPT-2, RoBERTa ve GPT-3 gibi son teknoloji Transformer tabanlı modellerin metin sınıflandırma görevlerindeki performansını değerlendirmektir. Bulgular, Naive Bayes'in geleneksel yöntemler arasında %65'lik bir doğruluk oranına ulaştığını gösterirken, GPT-3'ün onları %77 daha yüksek bir doğruluk ve F1 skoru ile aştığını ortaya koymaktadır. Bu sonuçlar, Transformer tabanlı modellerin metin sınıflandırma çabalarında önemli vaatler ve etkinliklerini vurgulamaktadır.

Anahtar Kelimeler: Derin öğrenme, makine öğrenimi, doğal dil işleme, metin sınıflandırma, dönüştürücü.

1. INTRODUCTION

The evolution of Artificial Intelligence (AI) has followed a notable trajectory, commencing with Alan Turing's seminal inquiry, "Can machines think?" and subsequently marked by John McCarthy's introduction of the term AI, catalyzing intensified research efforts in universities during the 1960s. Despite the challenges encountered during the AI Winter in the 1980s, global AI research persisted, culminating in significant breakthroughs in learning algorithms and image recognition technologies. Landmark events such as IBM's Deep Blue defeating a world champion chess player reignited interest in AI, prompting extensive investments by significant tech companies [1].

As digitalization advances rapidly in the information age, AI is assuming a pivotal role, much like other technologies, infiltrating various facets of human existence. AI, characterized by systems and machines endeavoring to replicate human intelligence through data analysis and interpretation, promises to enrich human experiences, boost productivity, and simplify daily life. Concurrently, as technology progresses, the cost of AI applications declines while their performance consistently ascends [2].

Machine Learning (ML), a popular concept in technology, has had broad usage for a considerable period. However, the decline in storage costs and the increase in computational power have accelerated the rise

**Corresponding Author*
e-posta : nazif.aydin@gazi.edu.tr

of Deep Learning (DL) in recent years. This trend allowed DL to gain more popularity in application and research areas within AI compared to ML. Especially in processing large datasets and solving complex problems, DL has garnered significant attention in this new era. The decreasing costs and increasing performance have notably enhanced the widespread usability of DL, significantly boosting interest compared to ML.

In recent years, notable strides in DL and Artificial Neural Networks (ANNs) have spurred remarkable advancements in NLP. Particularly, a rapid acceleration in DL-focused research and studies has been observed. One of the most remarkable advancements in this field is the continuously evolving transformer-based systems built upon DL language models [3]. These advancements in AI technologies significantly expand the horizons for research in NLP, considerably increasing interest in this domain.

In a study by Chen et al., a machine learning and deep learning model is proposed for text classification, utilizing joint training and an attention mechanism to improve the classification of both short and long texts [4]. In their study, Mokhamed and colleagues present a comparative analysis of machine learning (ML) and deep learning (DL) methods for emoji prediction from Arabic texts [5]. Another study by Chhabra et al., various machine learning and deep learning algorithms were used to classify Hindi news articles published in Hindi-language newspapers in India [6].A study aims to compare the effectiveness of DL methods, including Convolutional Neural Networks, with traditional ML techniques such as NB and SVM in document classification. The objective is to evaluate these contrasting approaches' accuracy, speed, and efficiency and discern their effectiveness across diverse scenarios [7]. Focusing on this objective, the study aims to ascertain the performance disparities among various learning methodologies employed in document classification, striving to pinpoint the most optimal techniques. In another study, the significance of text classification as a burgeoning research domain over the past decade, particularly with the advent of DL techniques, is underscored. Tracing the evolution of traditional and DL models in text classification, the article delves into an analysis of studies conducted in this field, examining distinct approaches, datasets utilized, models employed, and evaluation metrics applied. Moreover, it offers valuable insights into prospective avenues for research and sheds light on the persistent challenges within the domain [8].

Transformer-based models, such as those mentioned, have exhibited remarkable effectiveness in tasks like text classification when contrasted with conventional ML algorithms [9]. Their success in processing textual data surpasses previous methodologies, sparking substantial interest in AI. Integrating innovative approaches like transformer models alongside traditional methods signifies a pivotal juncture in AI advancement. These novel models have demonstrated superior performance in handling extensive text datasets, spanning domains like language processing, translation, and text comprehension, thus accelerating AI's penetration into diverse application domains and elevating its significance as a focal point [10].

This study delves into the comparison between next-generation transformer-based DL algorithms (GPT-2, GPT-3, BERT, DistilBERT, RoBERTa) and traditional text classification algorithms (DT, NB, RF, and SVM) within the education sector. The algorithms undergo classification tasks using a dataset encompassing six distinct categories: preliminary education, elementary education, intermediate education, secondary education, undergraduate studies, and postgraduate education. Performance assessment of each algorithm is conducted utilizing key parameters, including accuracy, F1 score, and recall, among others.

2. MATERIAL and METHOD

In this section, attention-based DL techniques, such as transformer-based language models and techniques, as well as terminology related to ML, will be examined.

2.1. Natural Language Processing

NLP is a crucial discipline bridging linguistics and computer science, attaining heightened relevance amidst the swift march of technological progress [11]. Research endeavors within this domain concentrate on the analysis, interpretation, and generation of data in natural language by computer systems [12]. Recent strides in DL and ANNs methodologies have notably propelled substantial advancements in NLP. Fig. 1 provides illustrative examples underscoring these noteworthy developments, accentuating the ongoing research and progress within this sphere.

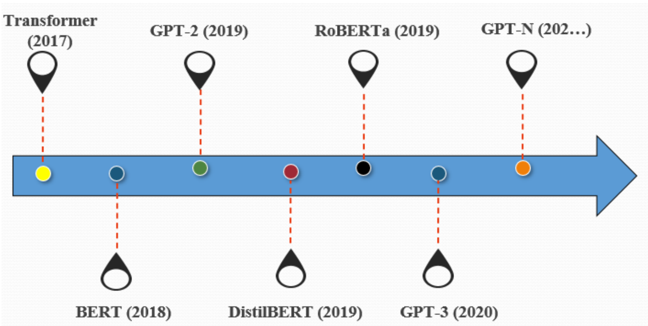

Figure 1. NLP key developments

The Transformer, a foundational framework for language models, has emerged as a potent AI model, demonstrating considerable prowess in handling sequential data. Among its derivatives, RoBERTa, a product of Facebook's ingenuity, leverages Transformer features to excel in text processing endeavors, showcasing remarkable performance. DistilBERT, a streamlined iteration of the BERT model, boasts enhanced speed and efficiency in information processing tasks. BERT, an offering from Google's research arsenal, stands out as a pre-trained language model, distinguished by its capacity to generate bidirectional language representations, proving effective across a spectrum of NLP tasks. On the other hand, GPT (Generative Pre-trained Transformer), an innovation from OpenAI, exhibits a remarkable ability to produce human-like texts, particularly excelling in large-scale text generation tasks. These versatile models can be pre-trained to address various natural language processing tasks and fine-tuned to cater to specific application requirements [13].

2.2. Transformer Model

Introduced by Google and researchers from the University of Toronto in 2017, the Transformer model represents a groundbreaking network architecture founded on self-attention mechanisms [14]. This model typically comprises two core components, as often depicted in Figure 2: an encoder and a decoder. The encoder, composed of N layers, processes input data into a coded representation via internal functions. Subsequently, the decoder layers undertake the task of decoding this encoded representation. Notably, the self-attention mechanism embedded within the Transformer model distinguishes itself by its capability to discern semantic relationships within input data, leveraging contextual information. This unique attribute empowers the model to deliver exceptional performance across a spectrum of NLP applications, spanning classification, question-answering, summarization, translation, and text generation [15].

The operational essence of the Transformer architecture, as depicted in Figure 2, can be succinctly summarized as follows: Initially, text inputs are transformed into numerical vectors during the input embedding phase. Subsequently, positional encoding ensures the ordered processing of input data. The multi-head attention mechanism discerns textual relationships, while residual connections facilitate inter-layer interaction, thus preventing loss of information. Layer normalization further bolsters model performance by standardizing weights and activations. The feed-forward neural network then processes the embedding vectors to yield the ultimate output. This architecture represents a significant leap in natural language processing, adept at discerning semantic relationships within textual data. The operational flow generally entails converting input embedding vectors and word-based representations into numerical form, ensuring sequential word positions through positional encoding via sine and cosine functions. The multi-head attention mechanism identifies and computes attention weights for pertinent words. The input vector undergoes normalization and adaptation to fit the decoder block. Similar processing occurs for the encoder's output in the decoder layer, culminating in the decoder's output obtained through linear function and softmax operations [16].

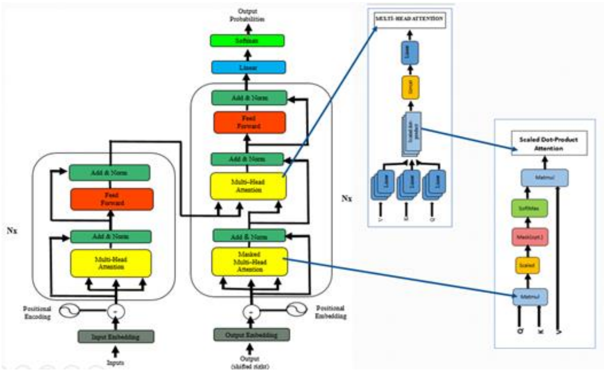

Figure 2. Model of transformer network architecture [14]

In recent years, a notable transformation has occurred within NLP, marked by a preference for the Transformer architecture over traditional methods such as LSTM and RNN. This newfound architecture distinguishes itself through its rapid processing capabilities, parallel operations, and adeptness at preserving contextual meanings over extended passages. Concurrently, a plethora of foundational Transformer models has emerged. Large Language Models(LLMs) like GPT, BERT, Hugging Face, and Turing NLG have found success across diverse domains. These models have been effectively utilized for tasks ranging from generating creative texts to furnishing accurate and efficient responses and facilitating text classification endeavors [17].

2.2.1. BERT

In 2018, Google introduced BERT (Bidirectional Encoder Representations from Transformers) through a seminal paper titled 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding' [18]. This groundbreaking language model marks a significant leap forward in natural language processing. Trained on an extensive corpus of text data, BERT has demonstrated exceptional prowess in grasping intricate word relationships and has proven highly effective across various language processing tasks, including text classification and precise predictions. The paper meticulously delineates BERT's pre-training methodology, employing a two-stage approach, and thoroughly elucidates how this model surpasses its predecessors in addressing natural language processing challenges. BERT is renowned for its capacity to comprehend contextual nuances within texts and its remarkable level of generalization, attained through training on vast datasets.

2.2.2. GPT2

GPT-2 is a language model developed by OpenAI and introduced in an article titled 'Language Models are Unsupervised Multitask Learners' [19]. Leveraging extensive training on large-scale text data, this model has showcased superior performance in NLP tasks compared to its predecessors. GPT-2 has garnered significant attention for its exceptional capabilities across various tasks, including language generation, text completion, translation, and text classification. The article provides a comprehensive exposition of the model's pre-training methodology and remarkable generalization abilities. At the heart of GPT-2 lies its remarkable language generation and comprehension proficiency, underpinned by its training on extensive text datasets.

2.2.3. DistilBERT

DistilBERT is a scaled-down language model developed by Hugging Face and introduced in an article titled 'DistilBERT, a distilled version of BERT: smaller, faster, cheaper, and lighter' [20], represents a compact language model developed by Hugging Face. Inspired by BERT (Bidirectional Encoder Representations from Transformers), DistilBERT aims to offer a smaller, swifter, and lighter alternative. By significantly reducing the size of pre-trained large language models, DistilBERT provides notable computational advantages, notably shortening training times. It has showcased effective performance across various NLP tasks such as text classification, sentiment analysis, and question-answering. The article delves extensively into the advantages of DistilBERT over BERT, detailing its training strategies and performance metrics.

2.2.4. RoBERTa

RoBERTa, introduced in a paper titled 'RoBERTa: A Robustly Optimized BERT Pretraining Approach' [21], represents a language model developed by Facebook AI. Building upon BERT's (Bidirectional Encoder Representations from Transformers) self-attention mechanism, RoBERTa undergoes more robust pre-training with a deeper and broader training set. The primary objective of RoBERTa is to bolster BERT's performance by implementing diverse features and techniques during the pre-training phase of language models. Moreover, it has been pre-trained utilizing extensive datasets, mainly geared towards enhancing performance in natural language processing tasks. The paper extensively elucidates RoBERTa's distinctive attributes, pre-training methodologies, and performance metrics.

2.2.5. GPT-3

GPT-3, presented in a paper titled 'Language Models are Few-Shot Learners,' is a highly extensive language model developed by OpenAI [22]. Distinguished by its substantially augmented scale compared to previous iterations, GPT-3 boasts a remarkable capability for few-shot learning, necessitating only a minimal number of examples for proficiency. With a staggering 175 billion parameters, this model demonstrates unparalleled performance across a plethora of tasks, including text generation, translation, question-answering, text classification, and numerous other language-processing endeavors. The paper extensively explores the implications of GPT-3's size and learning capabilities within the language processing domain, meticulously delineating the model's competencies and constraints.

2.3. Traditional ML Methods

Traditional machine and DL methods have played a significant role in NLP problems like text classification, showcasing various advantages and limitations [23]. These methods, including DT, NB, RF, and SVM are adept at constructing effective models that categorize text into specific classes based on word features present in the text. Naive Bayes, for instance, relies on the assumption of independence among word features and is often an initial choice for text classification tasks. Conversely, Decision Tree structures text data in a tree format to make decisions, while Random Forest amalgamates multiple decision trees to explore intricate relationships within the data. In contrast, SVM establishes separation boundaries among text classes, achieving superior text classification performance. However, it's worth noting that these traditional methods may face challenges in capturing nuanced semantic relationships and intricate textual structures. Notably, their efficacy might diminish when confronted with large, high-dimensional text datasets [24].

2.3.1. Decision tree

The paper titled 'Random Forests,' published by the renowned scientist Leo Breiman in 2001, introduced Random Forests, a significant expansion and advancement of the Decision Tree algorithm. This work is built upon the fundamental logic of the traditional Decision Tree algorithm, presenting a new algorithm that involves the collective construction of multiple decision trees and their collaboration. Like trees in a forest, this method comprises multiple decision trees working together. Each tree is trained with different features and subsets of data, and their results are aggregated to enable more robust and generalized predictions. Breiman's work marked a pivotal moment in developing models based on decision trees in ML and data mining [25].

2.3.2. Naive bayes

The Naive Bayes algorithm is a straightforward yet powerful probability-based classification technique rooted in Thomas Bayes' theorem. Leveraging this theorem's principles, the algorithm computes the probability of features within a dataset belonging to specific classes. Particularly notable for its efficacy in text classification, Naive Bayes is esteemed for its simplicity and swift execution. Numerous studies have delved into the Naive Bayes algorithm's intricacies, including the seminal book 'C4.5: Programs for Machine Learning,' by John R. Quinlan in 1993. This comprehensive resource offers in-depth insights into the characteristics and practical applications of decision trees and Naive Bayes classifiers. Quinlan's seminal work remains a cornerstone reference in exploring Naive Bayes and other ML methodologies.

2.3.3. Random forest

The Random Forest algorithm represents an ensemble learning technique comprising a collection of decision trees. This approach fosters collaboration among diverse decision trees, culminating in a more resilient and dependable classification model. Operating autonomously, each tree is trained on distinct subsets of the data, thereby promoting diversity within the ensemble. Subsequently, individual trees generate their predictions aggregated to yield the final prediction. The Random Forest algorithm has garnered considerable attention within the ML community and is frequently favored for its effectiveness in addressing classification and regression challenges. A seminal article by Leo Breiman in 2001 is a pivotal contribution, offering an extensive analysis of the Random Forest algorithm's foundational principles, applications, and efficacy.

2.3.4. SVM

The Support Vector Machine (SVM) is a formidable ML algorithm widely employed in classification and regression tasks. SVM creates a hyperplane to effectively segregate data into distinct classes, meticulously optimizing the separation of data points along this plane. At its core, SVM strives to identify the hyperplane that maximizes the margin—the distance between the plane and the data points—resulting in optimal classification. The seminal introduction of SVM was elucidated by Vladimir Vapnik and his team in their seminal work "The Nature of Statistical Learning Theory," published in 1995. This seminal publication offers a comprehensive exposition of the theoretical underpinnings of SVM and its efficacy in tackling classification challenges, delving into the intricate mathematical foundations of the algorithm. As a pivotal reference, the book serves as an indispensable resource for comprehending the operational principles of SVM and its performance in classification scenarios.

2.4. Experiment Evaluation

In preparation for this study, a dataset has been meticulously compiled by analyzing various documents, including laws, regulations, bylaws, directives, procedures, and principles utilized by the Ministry of National Education and the Council of Higher Education in student educational processes. This dataset encompasses a total of 476 samples, meticulously curated to capture the nuances of educational practices. The dataset has been meticulously categorized according to educational levels, encompassing preliminary education, elementary education, intermediate education, secondary education, undergraduate studies, and postgraduate education levels, as delineated in Table 1.

Table 1. Training dataset

Level	Education Category	Sample
1	preliminary education	32
2	elementary education	30
3	intermediate education	150
4	secondary education	135
5	undergraduate studies	40
6	postgraduat education	89
	Total	**476**

In comparing transformer-based innovative models and traditional ML algorithms, model performance is evaluated using various criteria and metrics. Among these metrics, key performance measures such as accuracy, recall, confusion matrix, F1-score, and precision are essential. The confusion matrix, for instance, provides a tabular representation that contrasts the classification model's predictions with the actual classes, partitioned into four main sections: False Negative (FN), True Positive (TP), True Negative (TN), and False Positive (FP). These terms are instrumental in analyzing the model's accuracy, identifying misclassifications, and assessing overall performance. These metrics play a pivotal role in elucidating the efficacy of Transformer models in text classification tasks. Each metric, offering distinct perspectives on the model's performance, delineates its strengths and weaknesses in greater detail. An objective and comprehensive assessment of the model's classification performance can be achieved by employing these metrics. These metrics and their corresponding values are outlined in Table 2.

Table 2. Confusion matrix

		Actual (Reference) Values	
		Positive	Negative
Predicted Values	Positive	TP	FP
	Negative	FN	TN

Below are the performance metrics obtained from the table along with their formulas:

2.4.1. Accuracy

Accuracy is a metric that denotes the proportion of correctly predicted instances relative to the total number of instances. It offers insight into the overall effectiveness of a model's predictions. The formalized definition of accuracy is represented by Equation (1) below.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \quad (1)$$

2.4.2. Precision

Precision quantifies the proportion of correctly predicted positive instances out of all instances predicted as positive. It serves to mitigate false positives, offering a measure of the model's precision in identifying positive cases. The formal definition of precision is depicted in Equation (2) below.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

2.4.3. Recall

Recall, also known as sensitivity or true positive rate, measures the proportion of true positives correctly identified by the model among all actual positive instances. Its objective is to minimize false negatives, providing insight into the model's ability to capture all positive cases. The formal definition of recall is illustrated in Equation (3) below.

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

2.4.4. F1 score

The F1 score is a harmonic mean of precision and recall, offering a balanced assessment of both metrics. It provides a consolidated measure of the model's performance, simultaneously considering precision and recall. The formal definition of the F1 score is depicted in Equation (4) below.

$$\text{F1Score} = \frac{2*(Precision*Recall)}{Precision+Recall} \tag{4}$$

These formulas constitute fundamental metrics employed for evaluating the performance of classification models. Precision, recall, and F1 scores hold particular significance as they offer a balanced assessment, considering false positives and false negatives. While accuracy provides an overall measure of correctness, it may not suffice, especially in scenarios involving imbalanced classification problems. Therefore, precision, recall, and F1 score play a crucial role in providing a more nuanced evaluation of the model's performance across different aspects of classification.

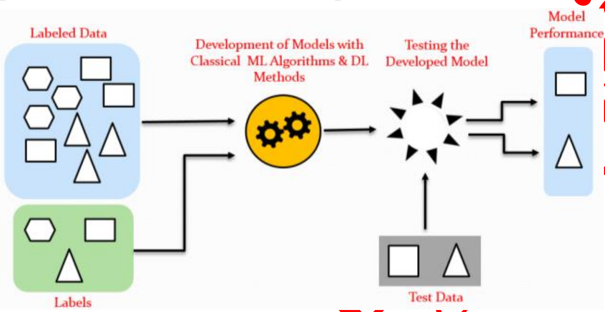

Figure 3. The overall workflow diagram and architecture of the study

The architecture implemented in this study, as shown in Figure 3, primarily revolves around utilizing a labeled dataset, which assumes a pivotal role in training both innovative models and traditional machine learning algorithms. This training process involves presenting data to the model and imparting it with correct labels, employing a blend of classical machine learning algorithms (referred to as model development) and applying deep learning techniques. Upon completing the training phase, the model undergoes evaluation using a distinct test dataset, which comprises data the model hasn't encountered during training and differs from the dataset utilized in the training phase. Subsequently, the model generates predictions on the input data within the test dataset, and these predictions are juxtaposed with the actual labels. This evaluation phase is essential for gauging the model's performance and ensuring accurate predictions on unseen data. It critically assesses the model's adaptability to the training data and its efficacy in handling novel data instances. Within the study's framework, both traditional methods and innovative models underwent training and testing using identical datasets. The resultant accuracy, precision, recall, and F1 score metrics are compiled and presented in Table 3 for comprehensive analysis and comparison.

Analyzing the data provided in the table enables a comprehensive comparison of the performance between traditional and innovative models:

- GPT-3 stands out with the highest accuracy and F1 score.
- Naive Bayes exhibits high performance compared to other models, demonstrating good precision and F1 score overall.
- RoBERTa demonstrates average performance, with a higher accuracy value than other models but moderate precision and recall values.
- GPT-2 and DistilBERT generally show lower performance, particularly with low precision and F1 score.
- BERT and Decision Tree models are among the ones with lower performance metrics

Table 3. Performance comparison

Model	**Accuracy**	**Precision**	**Recall**	**F1**
BERT	**0.50**	0.45	0.36	0.34
GPT-2	**0.44**	0.38	0.44	0.39
DistilBERT	**0.50**	0.28	0.34	0.28
RoBERTa	**0.56**	0.297	0.37	0.32
GPT-3	**0.77**	-	-	**0.77**
DT	**0.47**	0.43	**0.47**	0.44
NB	**0.65**	**0.65**	**0.65**	0.61
RF	**0.61**	0.60	**0.61**	0.58
SVM	**0.59**	0.57	0.54	0.55

While GPT-3 and Naive Bayes models exhibit higher performance, GPT-2 and DistilBERT show lower performance.

3. CONCLUSION

In today's era, easy access to data, high computational capabilities, and practical algorithms have amplified the significance of utilizing information processing for the benefit of humanity. In Industry 4.0, AI presents significant opportunities for enhancing productivity and quality of life. Mainly, advancements in NLP and DL have facilitated the development of models capable of conversing, writing, and even generating texts that resemble human-generated content.

This study assesses the efficacy of Transformer-based innovative models, leveraging attention-based DL architecture, in text classification endeavors juxtaposed with traditional ML algorithms. The experiments elucidated that traditional algorithms, including DT, NB, RF, and SVM attained a maximum accuracy rate of 65%. Conversely, among Transformer-based DL methodologies such as BERT, DistilBERT, GPT-2, GPT-3, and RoBERTa, notably, GPT-3 showcased a substantially higher accuracy rate and F1 score, surpassing other DL models and conventional ML algorithms by 77%. These findings underscore the superiority of Transformer models over traditional algorithms, particularly in achieving superior outcomes on extensive datasets and delivering elevated accuracy levels in text classification endeavors. Large Transformer models, particularly those rooted in GPT architecture, demonstrate an exceptional capacity for comprehensive learning on vast datasets and intricate textual structures. Such results accentuate the transformative potential of Transformer-based models in text classification, heralding promising prospects for future advancements in this domain.

The findings of this study indicate that Transformer-based models might offer a more effective alternative to traditional ML algorithms in text classification. Particularly in industrial applications, widespread use of these models, which work on large datasets and excel in DL capabilities, could provide better and more precise solutions to text classification problems. Consequently, future research might delve deeper into these innovative models' application and generalization capabilities across various industrial and academic domains.

ACKNOWLEDGEMENT

This study is supported by NVIDIA Corporation. All experimental studies were performed on a TITAN XP graphic card sdonated by NVIDIA. We would like to thank NVIDIA for support.

DECLARATION OF ETHICAL STANDARDS

The author(s) of this article declare that the materials and methods used in their study do not require approval from an ethics committee and/or any specific legal or private permissions.

AUTHORS' CONTRIBUTIONS

Adem TEKEREK: Perofrmed the experiments and analyse the results.

O. Ayhan ERDEM: Perofrmed the experiments and analyse the results.

Nazif AYDIN: Perofrmed the experiments and analyse the results and wrote the manuscript.

CONFLICT OF INTEREST

There is no conflict of interest in this study.

REFERENCES

[1] Göksu G., "Türkiye Yapay Zeka İnisiyatifi 1. Çalıştay Raporu", ***GelecekHane A.Ş.***, (2017).

[2] Hirvonen, N., Jylhä, V., Lao, Y., and Larsson, S., "Artificial intelligence in the information ecosystem: Affordances for everyday information seeking", **Journal of the Association for Information Science and Technology**, (2023).

[3] Aydın N. and Erdem O.A, "A research on the new generation artificial intelligence technology generative pretraining transformer 3" ***2022 3rd International Informatics and Software Engineering Conference (IISEC)***, (2022).

[4] Chen, H., Zhang, H., Yang, Y., and He, L.,"A text classification network model combining machine learning and deep learning", ***International Journal of Sensor Networks***, 44(3): 182-192, (2024).

[5] Mokhamed, T., Harous, S., Hussein, N., and Ismail, H., "Comparative analysis of deep learning and machine learning algorithms for emoji prediction from arabic text", ***Social Network Analysis and Mining***, 14(1),: 67, (2024).

[6] Chhabra, A., Arora, M., Sharma, A., Singh, H., Verma, S., Jain, R., ... and Kanavos, A., "Classifying hindi news using various machine learning and deep learning techniques", ***International Journal on Artificial Intelligence Tools***, (2023).

[7] Kamath, C. N., Bukhari, S. S., and Dengel, A. "Comparative study between traditional machine learning and deep learning approaches for text classification", ***In Proceedings of the ACM Symposium on Document Engineering 2018***, 1-11,(2018).

[8] Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., ... and He, L., "A survey on text classification: from traditional to deep learning." ***ACM Transactions on Intelligent Systems and Technology (TIST)***, 13(2): 1-41, (2022).

[9] Ikonomakis, M., Kotsiantis, S., and Tampakas, V., "Text classification using machine learning techniques", ***WSEAS Transactions on Computers***, 4(8): 966-974, (2005).

[10] Mukhamediev, R. I., Popova, Y., Kuchin, Y., Zaitseva, E., Kalimoldayev, A., Symagulov, A., ... and Yelis, M., "Review of artificial intelligence and machine learning technologies: classification, restrictions, opportunities and challenges", ***Mathematics***, 10(15):2552, (2022).

[11] Doszkocs, T. E., "Natural language processing in information retrieval", ***Journal of the American Society for Information Science***, 37(4): 191-196, (1986).

[12] Kuşçu, E., "Çeviride Yapay Zeka Uygulamalari", ***Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi***, 30: 45-58, (2015).

[13] Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., and Wang, Z., "ChatGPT and a new academic reality: artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing", ***Journal of the Association for Information Science and Technology***, 74(5): 570-581, (2023).

[14] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... and Polosukhin, I., "Attention is all you need", ***Advances in Neural Information Processing Systems***, (2017).

[15] Wang, Z., He, S., Xu, G., and Ren, M., "Will sentiment analysis need subculture? A new data augmentation approach", ***Journal of the Association for Information Science and Technology,*** 75(6): 655-670, (2024).

[16] Aydın, N., "A Review On The New Generation Technology Artificial Intelligence Generative Pre-Trained Transformer" ***Master's Thesis***, Gazi University, (2023).

[17] Amatriain, X., Sankar, A., Bing, J., Bodigutla, P. K., Hazen, T. J., and Kazi, M., "Transformer models: an introduction and catalog", ***arXiv preprint arXiv:2302.07730***., (2023).

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

[19] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I., "Language models are unsupervised multitask learners", ***OpenAI blog***, 1(8):9, (2019).

[20] Sanh, V., Debut, L., Chaumond, J., and Wolf, T., "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter", ***arXiv preprint arXiv:1910.01108.***, (2019).

[21] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... and Stoyanov, V., "Roberta: A robustly optimized bert pretraining approach", ***arXiv preprint arXiv:1907.11692.***, (2019).

[22] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... and Amodei, D., "Language models are few-shot learners", ***Advances in Neural Information Processing Systems***, 33: 1877-1901, (2020).

[23] Suissa, O., Elmalech, A., and Zhitomirsky-Geffet, M., "Text analysis using deep neural networks in digital humanities and information science", ***Journal of the Association for Information Science and Technology***, 73(2): 268-287, (2022).

[24] Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., and Brown, D., "Text classification algorithms: A survey", ***Information***, 10(4): 150, (2019).

[25] Breiman, L., "Random forests", ***Machine Learning***, 45: 5-32, (2001).