Classical and Deep Learning Approaches for Large-Scale Academic Text Classification

1 st Efe Mehmet Karabulut
Software Engineering
AVL Research and Engineering Istanbul, Türkiye
efe.karabulut@avl.com

2 st Irfan KOSESOY
Software Engineering
Kocaeli University Kocaeli, Turkey
irfan.kosesoy@kocaeli.edu.tr

Abstract **— This study aims to perform text classification on a large-scale dataset of academic abstracts. The dataset, specifically compiled for this research, consists of 121,000 academic abstracts from various disciplines, obtained through the Web of Science (WoS) portal, and serves as a unique resource. To predict the academic field to which each abstract belongs, both traditional and deep learning-based classification algorithms—such as Naive Bayes, Support Vector Machines (SVM), Random Forest, and BERT—were applied. Throughout the study, various combinations were explored by tuning vectorization methods, hyperparameters, and model architectures, allowing for a comprehensive comparative analysis of different approaches. As a result of extensive experimentation, the methods that yielded the highest accuracy, precision, recall, and F1 scores were identified and used to finalize the study. This strategy not only improved classification performance but also significantly enhanced the scientific originality and contribution of the research. The results demonstrate that the BERT model outperforms the other approaches in terms of classification accuracy, although it requires greater computational resources and longer processing times. Overall, this study provides a comparative evaluation of the effectiveness of different classification methods on large and balanced datasets, while also highlighting the practical potential of deep learning-based models.**

Keywords — Text classification, academic abstracts, Web of Science, machine learning, deep learning, BERT, Naive Bayes, Support Vector Machines, Random Forest, model comparison, natural language processing.

I. INTRODUCTION

The automatic classification of academic studies based on their content has become a critical research area in the era of big data, particularly for the management and accessibility of scientific knowledge. Today, hundreds of thousands of academic publications are produced annually, and [1]manually classifying such a large volume of content has become unsustainable in terms of human effort, time, and cost. With the increasing prevalence of interdisciplinary research, the need to categorize unstructured academic texts has grown significantly, making this task ever more important for researchers.

In this study, a large-scale dataset consisting of 121,000 academic abstracts from various disciplines was utilized. These abstracts were collected from the Web of Science (WoS) portal and represent recently published scholarly work. The dataset is a unique, up-to-date, and original resource specifically compiled for this research, and to the best of our knowledge, has not previously appeared in the literature. To perform the text classification task, both traditional and deep learning-based algorithms were employed, including Naive Bayes, Support Vector Machines (SVM), Random Forest, and BERT. The performance metrics of each model were carefully tuned to achieve optimal classification results, supported by different vectorization techniques and hyperparameter optimization strategies. The models were evaluated and compared, and combinations of the most successful approaches were implemented to enhance performance.

In recent years, machine learning and deep learning algorithms have emerged as powerful tools for analyzing large-scale datasets in text classification tasks. These algorithms play a significant role across various domains, including finance, healthcare, e-commerce, social media, and academia. Text classification serves as a fundamental component for

979-8-3315-9727-6/25/$31.00 ©2025 IEEE

applications such as sentiment analysis, spam detection, fraud detection, and information retrieval.

Research has shown that combining deep learning techniques with traditional machine learning methods can improve classification accuracy [1]. In the medical domain, such hybrid approaches have been employed for analyzing medical texts, revealing the strengths and limitations of various models [2, 12, 23]. Similarly, in the field of neuroinformatics, machine learning has introduced new methodologies for text classification and clustering using specialized datasets [9].

Multi-label classification has achieved significant progress, particularly in languages such as Bengali, through the use of ML-KNN and neural networks [7, 25]. Error log classification has increased the adoption of NLP techniques in industrial decision support systems [15, 24]. Sentiment analysis has shown promising potential in depression detection and psychological assessments [16, 11], and it is also widely used to analyze consumer reviews in the e-commerce domain [20].

Comparative studies have examined the most effective machine learning models for sentiment and intent classification tasks [22, 13]. Resume classification has been employed to optimize recruitment processes, with ongoing research aimed at improving classification accuracy [17].

While deep learning demonstrates superior performance in classifying large-scale text data, it often requires substantial computational resources [21]. Hybrid approaches have been shown to be effective in enhancing classification accuracy [6]. Additionally, pre-trained language models and transformer-based architectures continue to improve text classification performance [14]. Deep learning-based embedding techniques better capture contextual relationships between words, contributing further to classification quality [14].

II. METHODOLOGY AND MATERIALS

This section outlines the dataset used in the text classification process, the feature extraction techniques (TF-IDF, BERT), and the prediction methods (SVM, RF, Naïve Bayes, BERT) employed in the study. Notably, the BERT approach was utilized both for feature extraction and during the classification stage.

A. Dataset

In this study, a large-scale academic text dataset was used, compiled exclusively for this research from the Web of Science (WoS) database. This dataset is original, up-to-date, and—to the best of our knowledge—previously unused in the literature. It consists of 121,000 academic abstracts selected from 12 distinct disciplines, each representing a different scientific field. The dataset was carefully constructed with interdisciplinary diversity in mind and includes articles from fields such as engineering, medicine, social sciences, natural sciences, agriculture, environmental sciences, information technology, health sciences, energy, materials science, economics, and education.

To ensure statistical significance and balanced learning across all classes, an equal number of abstracts were selected from each discipline. Unlike datasets limited to titles or keywords, this collection is based on the full abstract texts of academic papers, providing a deeper and more content-rich foundation for classification tasks.This dataset is not available in any open-source repository and was developed exclusively for the purposes of this research. Considering its volume, disciplinary diversity, and recency, the dataset serves as a unique resource for studies in text classification and natural language processing (NLP). Following text preprocessing procedures, the dataset was partitioned into training and test subsets using an 80–20 split, enabling the training and evaluation of machine learning models. Numerical details regarding the distribution of classes are presented in Table 1.

Category	**Category-wise Distribution of the Dataset**			
	Total Articles	***Training (20%)***	***Test (80%)***	***Invalid Data***
Geography	11.651	2.331	9.320	0%
Astronomy and Astrophysics	10,801	2,161	8,640	0%
Computer Science and Artificial Intelligence	10,673	2,135	8,538	0%
Environmental Science	10,495	2,099	8,396	0%
Materials Science, Coatings and Films	10,479	2,096	8,383	0%
Chemistry	10,310	2,062	8,248	0%
Physics	9,696	1,940	7,756	0%
Music	9,417	1,884	7,533	0%
Economics and Business	9,266	1,854	7,412	0%
Neurosciences	9,129	1,826	7,303	0%
Biology	8,999	1,800	7,199	0%
Medicine, Research and Experimental	8,864	1,773	7,091	0%

Table I. Category-wise Distribution of the Dataset

B. Feature Extraction Methods

Feature extraction involves converting textual data into numerical representations that can be processed by machine learning models. This step is crucial for enabling the model to grasp the structure and meaning within the text. In this project, two different feature extraction methods were employed: TF-IDF and BERT. TF-IDF is a statistical approach based on the frequency and distinctiveness of words. Although it disregards contextual relationships, it was preferred due to its low computational cost and compatibility with classical algorithms such as SVM, Naive Bayes, and Random Forest. TF-IDF is a term frequency-based text vectorization technique where TF (Term Frequency) represents how often a word appears in a document, and IDF (Inverse Document Frequency) accounts for how common or rare the word is across multiple documents. The product of these two values determines the importance of a word within a specific document. Mathematically, the TF-IDF value of a term is calculated by Equation (1):

$$\text{TF-IDF}(\text{T}, \text{D}) = \text{TF}(\text{T}, \text{D}) \times \text{IDF}(\text{T}) \tag{1}$$

TF of a term in a document refers to the frequency of that term within the document. IDF of a term is calculated as the logarithm of the total number of documents divided by the number of documents containing the term. Here, the total number of documents is represented by N, and the number of documents containing the term is represented by DF.

BERT, on the other hand, is a deep learning-based language model that incorporates context by analyzing words bidirectionally, producing contextualized features. These features were directly evaluated using BERT's own classification layer. Applying both methods separately allowed for a comparative analysis of the impact of classical versus contextual representations on classification performance.

C. Prediction Methods

Both traditional machine learning algorithms and the deep learning-based BERT model were employed for the classification of academic texts. Naive Bayes is a probabilistic classifier widely favored in text classification due to its computational efficiency and simplicity. Random Forest enhances classification accuracy and mitigates overfitting by aggregating the predictions of multiple decision trees. Support Vector Machines (SVM) are effective in handling high-dimensional text data by determining optimal decision boundaries that maximize class separation. Furthermore, the BERT (Bidirectional Encoder Representations from Transformers) model leverages bidirectional contextual information to generate richer semantic representations. This pre-trained model was fine-tuned specifically for the classification task and demonstrated superior accuracy compared to conventional approaches.

D. Preprocessing, Vectorization, and Modeling

To prepare the dataset for use in the experimental studies and to ensure compatibility with machine learning model training, the data was cleaned, structured, and preprocessed. The preprocessing pipeline was designed to eliminate noise, standardize the text, and optimize feature extraction, thereby enhancing the performance and reliability of the classification models. The cleaned and structured dataset was then split into training and testing subsets, with 20% allocated for training and 80% for testing (see Table 1). This partitioning was performed using the train_test_split() function from the Scikit-Learn library, with a fixed random seed of 42 to ensure reproducibility of the results.Text normalization and inconsistency correction were addressed through a thorough cleaning process.

Initially, all special characters, punctuation marks, and non-alphanumeric symbols were removed using regular expressions (regex). This step was critical for eliminating noise that could negatively impact feature extraction. Subsequently, consecutive whitespace characters were replaced with a single space to prevent unnecessary token fragmentation. To guarantee consistent text representation and avoid the generation of redundant features, all text was converted to lowercase. This transformation mitigated the expansion of the feature space caused by case sensitivity, thus improving classification accuracy. For the classification of academic abstracts into their respective categories, three traditional machine learning algorithms—Naïve Bayes, Random Forest (RF), and Support Vector Machines (SVM)—were implemented. The textual data was converted into numerical representations using the TF-IDF vectorization technique. The vectorizer was configured with a maximum document frequency threshold of 75% to filter out overly common terms, and a combination of unigrams and bigrams was employed to capture contextual information. Once the feature vectors were generated, they were fed into the classification models, which were then trained using the labeled training data. The Naïve Bayes classifier utilized the multinomial variant, suitable for count-based text features.

The Random Forest classifier leveraged an ensemble of decision trees to improve prediction accuracy and reduce overfitting. The Support Vector Machines model was configured with a linear kernel, recognized as an effective approach for classifying high-dimensional text data.

E. Evaluation Metrics

To assess and compare the performance of the models used in this study, well-established evaluation metrics commonly adopted in machine learning research were employed. The primary purpose of these metrics is to measure the accuracy and effectiveness of model predictions and to provide an objective assessment of model performance. The following section provides a detailed explanation of these performance metrics, including their definitions, significance, and mathematical formulations.

Precision: Also known as the positive predictive value, precision measures the proportion of correctly classified positive instances among all instances predicted as positive. It quantifies the model's accuracy in identifying relevant examples and is particularly important in scenarios where false positives carry significant consequences. Mathematically, precision is defined as:

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

Here, TP (True Positives) represents the number of correctly classified positive instances, while FP (False Positives) denotes the number of negative instances that were incorrectly predicted as positive. A high precision value indicates that the model makes very few false positive errors.

Recall: Also known as the true positive rate or sensitivity, recall measures the proportion of correctly classified positive instances among all actual positive instances in the dataset. It is particularly crucial in applications where false negatives carry significant costs, such as medical diagnosis or fraud detection. Recall is mathematically formulated as follows:

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

Here, FN (False Negatives) represents the number of actual positive instances that were incorrectly classified as negative. A high recall value indicates that the model successfully identifies most of the positive cases while minimizing false negatives.

F1-Score: The F1-score is the harmonic mean of precision and recall, providing a single metric that balances the trade-off between the two. It is particularly useful when there is an imbalance in class distribution, as it prevents either precision or recall from being disproportionately favored. The F1-score is calculated as follows:

$$\text{F1 Score} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall}) \quad (4)$$

This metric is advantageous when both false positives and false negatives need to be minimized simultaneously. A high F1 score indicates that the model performs well in terms of both precision and recall, making it a reliable indicator of overall classification performance.

Accuracy: Accuracy is the most commonly used evaluation metric that measures the proportion of correctly classified samples (both positive and negative) out of the total number of samples in the dataset. It is calculated as follows:

$$\text{Accuracy} = (TP + TN) / (TP + FP + TN + FN) \quad (5)$$

Here, TN (True Negatives) represents the number of correctly classified negative samples. While accuracy provides a general measure of performance, it may not be a reliable metric in cases of class imbalance where one class significantly outnumbers the other. In such scenarios, a model can achieve high accuracy by simply predicting the majority class, rendering precision, recall, and the F1-score more informative alternatives. Collectively, these evaluation metrics offer a comprehensive assessment of a classification model's performance. Depending on the application and the relative importance of false positives and false negatives, different metrics may be prioritized to optimize the model for specific real-world use cases.

III. RESULTS

Following the training phase, a comprehensive performance evaluation was conducted to assess the efficacy of each classification model. Predictions were made on the test dataset, and a detailed classification report was generated for each category, encompassing key metrics such as accuracy, precision, recall, and F1-score. This evaluation facilitated a comparative analysis of the classifiers, providing insights into their discriminative capabilities across the various publication categories.

Model	**Comparison of Models**			
	Precision	***Recall***	***F1-Score***	***Accuracy***
Naive Bayes	0.82	0.79	0.79	80%
Random Forest	0.78	0.77	0.77	78%
SVM	0.86	0.86	0.86	86%
BERT	0.87	0.87	0.87	87%

Table II. Experimental Results

The experimental results presented in Table II provide a significant comparison for evaluating the classification performance of machine learning and deep learning-based models.

When the overall accuracy rates are considered, the BERT model demonstrated the highest classification performance with an accuracy of 87%. This outcome highlights the superiority of transformer-based models in capturing contextual meaning. In particular, BERT excels by learning long-range semantic relationships within texts, outperforming traditional methods in this regard.

The experimental results presented in Table II offer a significant comparison for evaluating the classification performance of machine learning and deep learning-based models.

When overall accuracy rates are considered, the BERT model demonstrated the highest classification performance with an accuracy of 87%. This outcome underscores the superiority of transformer-based models in capturing contextual meaning. Specifically, BERT excels at learning long-range semantic relationships within texts, enabling it to delineate clearer boundaries between classes that traditional methods struggle to differentiate. The SVM model closely followed BERT with an accuracy of 86%, making it the most successful traditional approach. Notably, SVM showed impressive results in technical and descriptive categories such as "Materials Science" and "Physics." This can be attributed to SVM's ability to construct effective decision boundaries in high-dimensional vector spaces.

Random Forest, on the other hand, exhibited relatively lower performance with an accuracy of 78%. As a tree-based method, it faced challenges in capturing contextual transitions in texts; while it produced high precision values in some classes, its recall rates remained limited. The Naive Bayes algorithm also achieved a similar accuracy level of around 80%, with lower performance observed in certain disciplines, likely due to the assumption of feature independence not holding in complex academic texts.

Performance differences among the models were evident not only in accuracy but also through class-wise precision, recall, and F1-scores. Across all models, categories such as “Biology” and “Medical Research and Experimental Studies” yielded comparatively lower results, possibly due to the overlapping content with other disciplines. Even in these challenging categories, BERT delivered more balanced outcomes compared to traditional methods. Furthermore, in conceptually distinct areas like “Music,” all models achieved high success rates, with BERT attaining an almost perfect classification accuracy of 99%.

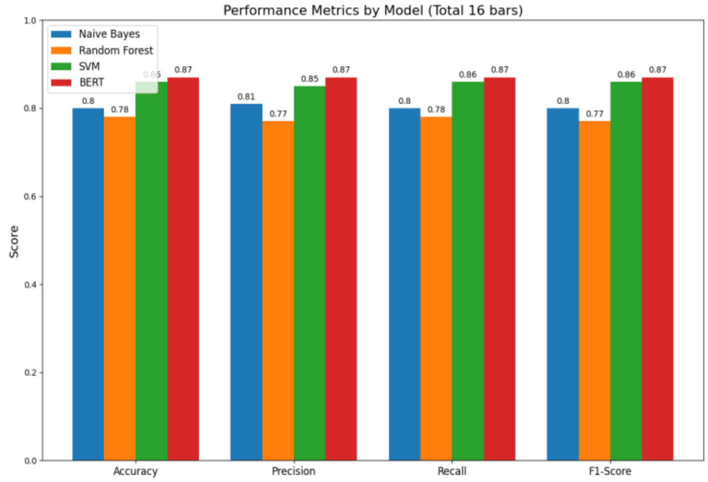

Fig. I. Visual representation of all the results shown in the table

In light of these findings, this study demonstrates that deep learning-based approaches, particularly transformer architectures like BERT, offer a substantial advantage in tasks demanding nuanced contextual analysis, such as the classification of academic texts. The superior performance of the BERT model, which achieved the highest overall accuracy on our bespoke, large-scale dataset, is particularly noteworthy. It underscores the effectiveness of these advanced methods in capturing long-range dependencies and intricate semantic relationships within such complex data. However, our findings also reveal that optimized traditional methods, such as the Support Vector Machine (SVM), remain highly competitive and retain their applicability in scenarios where computational cost and hardware constraints are critical considerations.

A key strength of this research lies in its methodological originality and rigor, which extends beyond the straightforward application of known techniques. While we utilized some of the most prominent methods from the current literature, our approach was distinctly unique. The classification process was meticulously executed on a large-scale dataset of 121,000 academic abstracts, which was uniquely compiled and curated for this study. The core of our contribution stems from the extensive and systematic experimentation with numerous combinations of data preprocessing, vectorization, and model training parameters. This comprehensive combinatorial approach, rather than a single-model focus, allowed us to identify the most successful and stable functional combinations, leading to the robust results presented. The detailed comparative analysis went beyond the core models, including an evaluation of various BERT derivatives (e.g., ALBERT, DistilBERT, SciBERT) and an exploration of diverse language processing techniques. Furthermore, the application of robust data splitting approaches like K-Fold cross-validation ensured the statistical reliability and generalizability of our findings. The comprehensive assessment also addressed the practical implications of these methods, detailing the rationale behind processor selection and providing a thorough analysis of the computational costs associated with each approach. Collectively, the combination of a unique, large-scale dataset and a highly original, experimental methodology clearly showcases the distinctive approach of this study, offering invaluable guidance for researchers and practitioners in method selection by highlighting the comparative strengths, limitations, and practical trade-offs of modern text classification techniques.

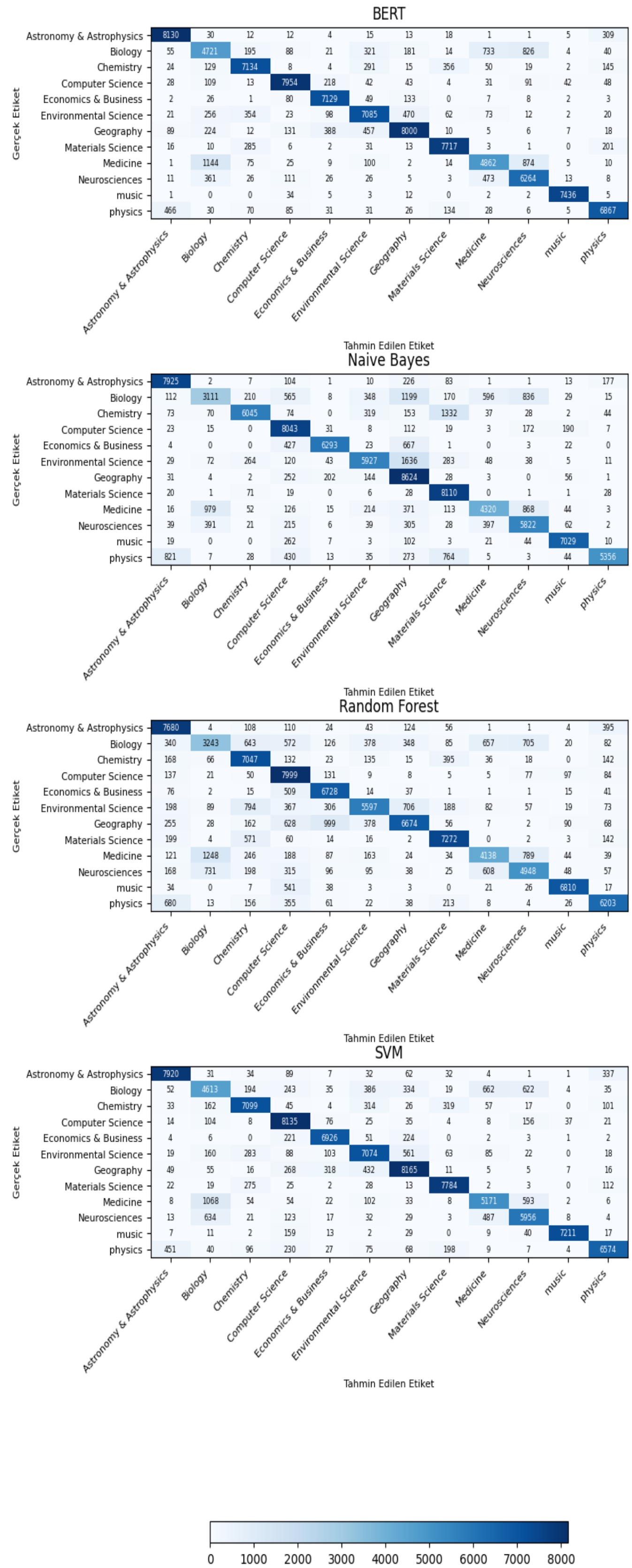

Fig. II. Confusion Matrix Heatmap of the Models

IV. Conclusion

In this study, we aimed to classify 121,000 scientific abstracts obtained from the Web of Science into 12 specific scientific categories, and we conducted a comprehensive comparison of different machine learning (ML) and deep learning (DL) models. The overall structure of the research is based on a stepwise improvement strategy designed to achieve more advanced systems in terms of both performance and efficiency.

During the model evaluation phase, rather than using k-fold cross-validation as the primary method, we chose an 80–20 train-test split strategy to provide a more controlled and stable metric measurement environment. This decision ensured fair and consistent comparisons among the Naive Bayes, Random Forest, SVM, and BERT models in the core experiment. Although k-fold cross-validation can be useful for assessing generalization performance, its inherent randomness can yield different scores even when running the same code multiple times. To minimize such variability and uncertainty in our performance metrics, all baseline comparisons were completed using a fixed-data split.

The results obtained are as follows:

- Naive Bayes: Completed training in 11.12 seconds and prediction in 21.62 seconds, achieving a weighted F1 score of 75.2%.
- Random Forest: Required 636.49 seconds for training and 42.18 seconds for prediction, with an F1 score of 77.6%.
- SVM: Incurred the highest computational cost—1,342.89 seconds for training and 3,148.52 seconds for prediction—while achieving an F1 score of 86.5%.

When examining transformer-based approaches:

- DistilBERT: Achieved 89.6% F1 after 1 hour and 12 minutes of training as a lightweight competitive model.
- ALBERT: Delivered an 88.9% F1 score in 1 hour and 26 minutes of training, demonstrating lower GPU and memory usage thanks to parameter sharing.
- BERT: Obtained a 87.0% F1 score in a single 2-hour and 47-minute training session, representing the highest baseline performance.
- 5-Fold K-Fold BERT: With 5-fold cross-validation, each fold ran for an average of 1 hour and 12 minutes, totaling approximately 6 hours and 5 minutes of training and yielding a 90.0% F1 score.
- SciBERT (80–20 split): Recorded an 88.0% weighted F1 score after approximately 2 hours and 10 minutes of training on the same data split.
- 5-Fold K-Fold SciBERT: Demonstrated the highest performance and stability, achieving a 92.0% weighted F1 score after a cumulative ~7.5 hours of training.

Due to the large dataset size, all transformer-based experiments were conducted by renting NVIDIA A100 GPUs via Google Colab Pro+; otherwise, compilation and training of each model would have taken days on standard hardware.

In the initial stages, we also tested classical word embedding techniques such as Word2Vec, FastText, and GloVe in combination with SVM, Random Forest, and Naive Bayes. However, their low accuracy led us to adopt TfidfVectorizer as the most stable feature extraction method for these traditional classifiers.

Error analysis between the 12 categories revealed that “Medicine, Research & Experimental” and “Neurosciences” were the most frequently confused pair, primarily because these two fields share highly similar conceptual terminology. For example, the abstract “Huntington disease (HD) is one of at least nine polyglutamine disorders caused by a CAG expansion in the coding region of a disease-causing gene. These disorders are characterized by selective neuronal degeneration particularly in the striatum and cerebral cortex, leading to progressive motor dysfunction and cognitive decline.” was misclassified as “Medicine” by the baseline BERT model but correctly classified as “Neurosciences” by SciBERT.

In summary, the fixed 80–20 split comparisons established each model’s baseline performance, while k-fold validation and domain-adaptive models provided further improvements in accuracy and robustness. Lightweight models like DistilBERT and ALBERT offered balanced solutions under resource constraints, whereas large-scale language models such as BERT and SciBERT delivered superior performance and flexibility. This study reaffirms the importance of domain-specific embeddings and appropriate validation strategies in scientific text classification. Future work may further enhance classification accuracy by incorporating additional metadata (e.g., journal name, keywords, citation counts).

V. DATA AVAILABILITY

Appendix-1 and Appendix-2 contains the dataset used in this study.

Appendix-1:
https://github.com/efemehmetkarabulut/Academic-Abstract-Dataset-for-Large-Scale-Academic-Text-Classification/blob/306580dff1a6c7eaf63e08bc645ade877dcee274/bbc_data_Format_151K%201.7z

Appendix-2:
https://github.com/efemehmetkarabulut/Academic-Abstract-Dataset-for-Large-Scale-Academic-Text-Classification/blob/306580dff1a6c7eaf63e08bc645ade877dcee274/bbc_data_Format_151K%202.7z

VI. REFERENCES

[1] M. Bounabi, K. E. Moutaouakil and K. Satori, "Neural Embedding & Hybrid ML Models for Text Classification," *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Meknes, Morocco, 2020, pp. 1–6, doi: 10.1109/IRASET48871.2020.9092230.

[2] K. M. Chaitrashree, T. N. Sneha, S. R. Tanushree, G. R. Usha and T. C. Pramod, "Unstructured Medical Text Classification using Machine Learning and Deep Learning Approaches," *2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, Bangalore, India, 2021, pp. 446–451, doi: 10.1109/RTEICT52294.2021.9573667.

[3] P. Shiguihara and L. Berton, "Exploring Deep Neural Networks and Decision Tree for Spanish Text Classification," *2022 IEEE XXIX International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, Lima, Peru, 2022, pp. 1–5, doi: 10.1109/INTERCON55795.2022.9870087.

[4] A. Chahar, N. Patil, D. Walunj, S. R. T, R. Shah and H. Saratkar, "An Indispensable Contemplation on Natural Language Processing Using Ensemble Techniques for Text Classification," *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2022, pp. 1–7, doi: 10.1109/ICACCS54159.2022.9785015.

[5] W. Akanda and A. Uddin, "Multi-Label Bengali Article Classification using ML-KNN Algorithm and Neural Network," *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 2021, pp. 1–6, doi: 10.1109/ICICT4SD50815.2021.9396882.

[6] D. Patil, R. Lokare and S. Patil, "An Overview of Text Representation Techniques in Text Classification using Deep Learning Models," *2022 3rd International Conference for Emerging Technology (INCET)*, 2022, pp. 1–6, doi: 10.1109/INCET54531.2022.9825389.

[7] N. Melethadathil, P. Chellaiah, B. Nair and S. Diwakar, "Classification and Clustering for Neuroinformatics: Assessing the Efficacy on Reverse-Mapped NeuroNLP Data Using Standard ML Techniques," *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Kochi, India, 2015, pp. 918–924, doi: 10.1109/ICACCI.2015.7275751.

[8] S. Chowdhury and M. P. Schoen, "Research Paper Classification using Supervised Machine Learning Techniques," *2020 Intermountain Engineering, Technology and Computing (IETC)*, Orem, UT, USA, 2020, pp. 1–6, doi: 10.1109/IETC47856.2020.9249211.

[9] M. Shafiq, X. Yu, A. A. Laghari, L. Yao, N. K. Karn and F. Abdessamia, "WeChat Text and Picture Messages Service Flow Traffic Classification Using Machine Learning Technique," *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, Sydney, NSW, Australia, 2016, pp. 1286–1293, doi: 10.1109/HPCC-SmartCity-DSS.2016.0019.

[10] I. Akhadam and H. Ayyad, "Enhancing Arabic Text Classification: A Comparative Study of Machine Learning and Deep Learning Approaches," *2024 IEEE 12th International Symposium on Signal, Image, Video and Communications (ISIVC)*, Marrakech, Morocco, 2023, pp. 1–6, doi: 10.1109/ISIVC61350.2024.10577929.

[11] T. M. Midhan, P. Selvaraj, M. H. K. Raju, M. B. P. Reddy and T. Bhaskar, "Classification of Mental Health and Emotion of Human from Text using Machine Learning Approaches," *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, 2023, pp. 1–6, doi: 10.1109/ISCON57294.2023.10111973.

[12] S. Mondal, A. K. Barman, S. Basumatary, M. Barman, C. Rai and A. Nag, "Cancer Text Article Categorization and Prediction Model Based on Machine Learning Approach," *2023 IEEE 3rd Mysore Sub Section International Conference (MysuruCon)*, 2023, pp. 1–6, doi: 10.1109/MysuruCon59703.2023.10397005.

[13] A. Muhammad, S. Defit and G. W. N. Cahyo, "Comparison of Machine Learning Text Classification for Intent Sentiment Analysis," *2023 IEEE 7th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2023, pp. 1–6, doi: 10.1109/ICITISEE58992.2023.10405200.

[14] C. M. Suneera and J. Prakash, "Performance Analysis of Machine Learning and Deep Learning Models for Text Classification," *2020 IEEE 17th India Council International Conference (INDICON)*, 2020, pp. 1–6, doi: 10.1109/INDICON49873.2020.9342208.

[15] D.-N. C. A. Chidinma, B. K. Mishra and W. Sayers, "Fault Log Text Classification Using Natural Language Processing And Machine Learning For Decision Support," *2022 14th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, 2022, pp. 1–6, doi: 10.1109/SKIMA57145.2022.10029587.

[16] M. M. Alnaddaf and M. S. Başarslan, "Sentiment Analysis Using Various Machine Learning Techniques on Depression Review Data," *2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2024, pp. 1–6, doi: 10.1109/IDAP64064.2024.10710889.

[17] B. Surendiran, T. Paturu, H. V. Chirumamilla and M. N. R. Reddy, "Text Classification Techniques Using Word Embeddings and Deep Learning Models," *2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT)*, 2023, pp. 1–6, doi: 10.1109/IConSCEPT57958.2023.10169907.

[18] M. Aamir, L. J. and Sweety, "A Comparative Study of ML and DL Approaches for Twitter Sentiment Classification," *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)*, 2024, pp. 1–6, doi: 10.1109/ICEECT61758.2024.10738938.

[19] S. T. Selvi, P. Karthikeyan, A. Vincent, V. Abinaya, G. Neeraja and R. Deepika, "Text Mining Based Classification of Research Abstracts using Clustering and Ensemble Techniques," *2016 Eighth International Conference on Advanced Computing (ICoAC)*, 2016, pp. 1–6, doi: 10.1109/ICoAC.2017.7951736.

[20] P. Kathuria, P. Sethi and R. Negi, "Sentiment Analysis on E-commerce Reviews and Ratings Using ML & NLP Models to Understand Consumer Behavior," *2022 International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC)*, 2022, pp. 1–6, doi: 10.1109/ICMACC54824.2022.10093674.

[21] S. Rehman, A. Irtaza, M. Nawaz and H. Kibriya, "Text Document Classification Using Deep Learning Techniques," *2022 International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)*, 2022, pp. 1–6, doi: 10.1109/ETECTE55893.2022.10007316.

[22] A. Bhalla, "Comparative Analysis of Text Classification using SVM, Naïve Bayes, and KNN Models," *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2023, pp. 1–6, doi: 10.1109/ICIRCA57980.2023.10220728.

[23] İ Kösesoy, M Gök, T Kahveci " Prediction of host-pathogen protein interactions by extended network model " Turkish Journal of Biology, 2021, pp. 138–148, doi:10.3906/biy-2009-4.

[24] İ Kösesoy, "Prediction of POMA-G Score from Spatiotemporal Gait Parameters," Dicle University Faculty of Engineering Engineering Journal, 2023, pp. 555–561, Doi: 10.24012/dumf.1326557

[25] İ Kösesoy, C Öz, " Acquiring Kinematics of Lower extremity with Kinect," European Journal of Science and Technology, 2021, pp. 92–100, DOI: 10.31590/ejosat.1041675