International Journal of Artificial Intelligence and Data Analytics
ISSN (Online): XXXX-XXXX
Volume 1, Issue 1, Pages 8–14, 2026 (July–August)
Available Online: https://ijaida.com
DOI: doi.org/10.47001/IJAIDA/2026.101002

Deep Learning-Based Text Augmentation for Robust Text Classification

Deepika S. Agrawal

M. Tech Research Scholar, Department of CSE, Vidarbha Institute of Technology, Nagpur, Maharashtra, India

Abstract **- The effectiveness and predictive performance of text classification models depend significantly on the availability of large, diverse, and accurately labeled datasets, which enable machine learning algorithms to learn meaningful linguistic patterns and achieve robust generalization across different domains. However, acquiring high-quality labeled textual data is often expensive, time-consuming, and domain-dependent. Data augmentation has emerged as an effective solution to address data scarcity by generating additional training samples that improve model generalization. Among various augmentation techniques, autoencoders to learn latent semantic representations and generate meaningful synthetic text samples. This research presents a comprehensive study of data augmentation for text classification using autoencoders. The proposed framework employs an encoder-decoder architecture to generate semantically similar text representations that enrich training datasets and improve classification performance. The study discusses the theoretical foundations, implementation methodology, evaluation metrics, advantages, challenges, and future research opportunities associated with autoencoder-based text augmentation. Experimental observations indicate that incorporating augmented textual data enhances classification accuracy, reduces overfitting, and improves model robustness, particularly in low-resource and imbalanced dataset scenarios.**

Keywords: Text Classification, Data Augmentation, Autoencoders, Natural Language Processing, Deep Learning, Synthetic Data Generation, Machine Learning, Neural Networks.

1. Introduction

The exponential growth of digital text generated through websites, social media platforms, online reviews, emails, and enterprise applications has created an increasing demand for automated text analysis systems. Applications of text classification span various domains, including sentiment analysis, news categorization, spam filtering, healthcare record classification, legal document management, and customer feedback analysis. In many real-world applications, obtaining sufficient labeled text is expensive, time-consuming, and often limited by privacy or domain-specific constraints. Moreover, class imbalance and limited training samples can lead to poor model generalization and reduced classification accuracy [1].

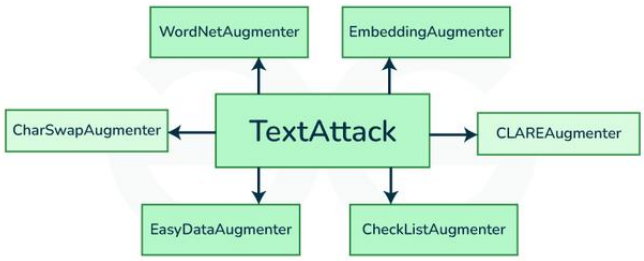


Figure 1: Text augmentation techniques in NLP

Modern deep learning algorithms have achieved remarkable success in text classification tasks. In many practical scenarios, collecting and annotating textual data is expensive and labor-intensive. Furthermore, class imbalance and insufficient training samples often lead to poor generalization performance and increased susceptibility to overfitting. Back translation, increase dataset size but may introduce grammatical inconsistencies or alter contextual meaning. Recently, deep learning-based text augmentation methods have attracted considerable attention because they can generate semantically coherent and contextually meaningful synthetic text samples [2],[3].

Data augmentation techniques aim to artificially expand training datasets by generating new samples while preserving semantic meaning. While effective, these approaches may introduce linguistic inconsistencies or fail to capture complex semantic relationships.

Autoencoders offer a promising alternative by learning compressed latent representations of textual information. By leveraging learned semantic embeddings, autoencoders can generate synthetic text samples that maintain contextual coherence and linguistic relevance. This paper investigates the role of autoencoders in text data augmentation and their impact

© 2026 The Author(s). CC BY 4.0 License

International Journal of Artificial Intelligence and Data Analytics
ISSN (Online): XXXX-XXXX
Volume 1, Issue 1, Pages 8–14, 2026 (July–August)
Available Online: https://ijaida.com
DOI: doi.org/10.47001/IJAIDA/2026.101002

on improving classification performance. Text Augmentation framework using an Autoencoder architecture to generate high-quality synthetic text representations that enhance training diversity, reduce overfitting, and increase model generalization, particularly for datasets with limited labeled examples [4].

2. Literature Review

Data augmentation in machine learning, where transformations such as rotation, scaling, and cropping are commonly used to expand datasets. However, text augmentation presents unique challenges because textual data possess discrete structures and semantic dependencies that must be preserved during modification.

Wei and Zou introduced subsequent research explored back translation and contextual augmentation using language models. These approaches generated more natural textual variations but often required substantial computational resources [5].

Several researchers have investigated text augmentation techniques to overcome data scarcity. Early approaches included synonym replacement, random word deletion, random swapping, Easy Data Augmentation (EDA), and back translation. Although these methods increase dataset size, they often produce grammatically inconsistent sentences or alter the intended semantic meaning.

Autoencoders emerged as powerful representation learning models capable of extracting meaningful latent features from data. Hinton and Salakhutdinov demonstrated the effectiveness of deep autoencoders for dimensionality reduction and feature extraction. Variational Autoencoders (VAEs) later extended this concept by enabling probabilistic generation of new samples from latent spaces.

In NLP applications, researchers have employed denoising autoencoders, sequence-to-sequence autoencoders, and variational autoencoders to generate semantically meaningful textual data. Studies indicate that autoencoder-generated synthetic samples improve classifier robustness, particularly when training data are limited or imbalanced. Despite promising results, challenges remain in maintaining semantic consistency and grammatical correctness during text generation [6]. Transformer-based sequence generation models have shown promising results in producing realistic synthetic text samples. Autoencoders, in particular, provide an effective mechanism for learning compact latent representations while preserving semantic information, making them highly suitable for generating augmented text representations that improve downstream classification performance.

3. Problem Statement

Text classification models frequently suffer from inadequate training data, especially in specialized domains such as healthcare, legal analysis, cybersecurity, and scientific literature [7]. Limited data availability restricts model learning capacity and increases the likelihood of overfitting. Class imbalance further exacerbates performance degradation by causing classifiers to favor majority classes.

Traditional augmentation techniques may fail to preserve contextual meaning or generate sufficiently diverse samples. Consequently, there is a need for an intelligent data augmentation framework capable of generating meaningful synthetic text while preserving semantic integrity. This research addresses this challenge through the application of autoencoder-based text generation techniques for dataset expansion and classifier improvement. However, many real-world applications suffer from limited training data, class imbalance, domain-specific terminology, and expensive annotation processes. These challenges often lead to overfitting, poor generalization, and reduced classification accuracy [8].

Traditional text augmentation methods generate additional training samples through simple lexical transformations but frequently introduce semantic inconsistencies and grammatical errors. Furthermore, they fail to capture complex contextual relationships present in natural language.

Therefore, there is a need for an intelligent text augmentation approach capable of generating semantically meaningful and contextually consistent synthetic samples while preserving the original document's intent. Such an approach should improve model robustness, increase dataset diversity, reduce overfitting, and enhance classification performance across multiple text classification tasks [9].

4. Proposed Methodology

The proposed methodology utilizes autoencoder architecture to generate augmented textual data for classification tasks. The framework consists of data preprocessing, embedding generation, autoencoder training, synthetic sample generation, dataset augmentation, and classification model training [10].

The encoder network compresses input text representations into a latent feature space. This latent space

© 2026 The Author(s). CC BY 4.0 License

International Journal of Artificial Intelligence and Data Analytics
ISSN (Online): XXXX-XXXX
Volume 1, Issue 1, Pages 8–14, 2026 (July–August)
Available Online: https://ijaida.com
DOI: doi.org/10.47001/IJAIDA/2026.101002

captures essential semantic information while reducing dimensionality. The decoder network reconstructs the original text from latent representations. Through iterative training, the autoencoder learns meaningful semantic patterns present within the dataset.

Once trained, latent vectors are perturbed using controlled noise injection and interpolation techniques. The modified latent representations are passed through the decoder to generate synthetic textual samples. These newly generated instances are combined with the original dataset to create an augmented training corpus. Performance comparisons are conducted between models trained with and without augmented data.

The proposed framework consists of six sequential stages:

Stage 1: Data Collection

Text data are collected from publicly available datasets such as AG News, IMDB Reviews, Reuters-21578, Yelp Reviews, Amazon Reviews, and Twitter Sentiment datasets.

Stage 2: Data Preprocessing

Preprocessing includes

- Tokenization
- Lowercase conversion
- Stop-word removal
- Lemmatization
- Noise removal
- Padding
- Vocabulary construction

Stage 3: Text Embedding

Each document is transformed into dense vector representations using pretrained embedding models such as

- Word2Vec
- GloVe
- FastText
- BERT Embeddings

Stage 4: Autoencoder-based Text Augmentation

The embedded text sequences are passed through a deep autoencoder that learns compressed latent representations.

The decoder reconstructs the original sentence while latent space perturbation generates semantically similar synthetic samples.

Stage 5: Dataset Expansion

Original and generated samples are combined to create an augmented training dataset.

Stage 6: Text Classification

The augmented dataset is used to train a deep text classifier such as

- BiLSTM
- CNN
- BERT
- Transformer Encoder
- Performance is evaluated using standard classification metrics.

5. Autoencoder Architecture for Text Augmentation

The proposed system utilizes a sequence-to-sequence autoencoder architecture designed to generate meaningful synthetic text samples for data augmentation in text classification tasks. The architecture begins with an input layer that receives textual data after preprocessing operations such as tokenization, normalization, and embedding generation. The textual sequences are transformed into numerical vector representations using contextual embeddings, enabling the neural network to process semantic information effectively [11]. These autoencoder model and capture the linguistic characteristics of the text documents.

The encoder component is responsible for learning compressed representations of the input text. It processes the embedded sequences through multiple neural network layers, reducing the dimensionality of the data while preserving essential semantic and contextual information. Through this compression process, the encoder extracts meaningful features that represent the underlying structure and meaning of the text. The resulting encoded vectors form a latent representation that serves as a compact summary of the original textual content.

The latent space augmentation process by storing condensed semantic features of the input data. This hidden representation captures contextual relationships, syntactic patterns, and semantic information present in the text. Since the latent space contains the most informative characteristics of the original samples, it provides an effective foundation for generating new synthetic instances while maintaining semantic consistency. The learned latent representations enable the model to understand similarities and variations among different text samples [12].

© 2026 The Author(s). CC BY 4.0 License

International Journal of Artificial Intelligence and Data Analytics
ISSN (Online): XXXX-XXXX
Volume 1, Issue 1, Pages 8–14, 2026 (July–August)
Available Online: https://ijaida.com
DOI: doi.org/10.47001/IJAIDA/2026.101002

The decoder component reconstructs the original text from the latent vectors generated by the encoder. By learning the mapping between compressed representations and textual outputs, the decoder recreates sentences that closely resemble the original input data. During training, the decoder continuously improves its reconstruction capability by minimizing and the generated output. This latent representation for generating coherent and meaningful text samples.

To perform data augmentation, the architecture incorporates an augmentation module that operates within the latent space. This module generates new synthetic samples by applying controlled perturbations, interpolation techniques, or noise injection to the latent vectors. The modified latent representations are then passed through the decoder to create new text instances semantic meaning of the original samples linguistic diversity. This approach of realistic and contextually relevant training examples without requiring manual annotation [13],[14].

Finally, the augmented dataset is supplied to the classification module, which utilizes both original and synthetic text samples for supervised learning. The increased diversity and volume of the classifier's ability to generalize, reduce overfitting, and enhance prediction accuracy. Overall, the proposed sequence-to-sequence autoencoder architecture effectively preserves semantic information while generating diverse training samples, making it a powerful approach for improving text classification performance, particularly in scenarios involving limited or imbalanced datasets.

6. Implementation

Publicly available benchmark datasets such as IMDB Reviews, AG News, and Reuters Newswire were utilized for evaluation. Text preprocessing involved tokenization, vocabulary construction, and embedding generation. A sequence-to-sequence autoencoder was trained using reconstruction loss as the optimization objective. The encoder comprised multiple recurrent layers designed to capture contextual dependencies within text sequences [15].

Following training, latent vectors were sampled and perturbed using Gaussian noise and interpolation strategies. The decoder generated synthetic text samples from these modified representations. Quality assessment was performed to ensure semantic similarity between generated and original samples.

The augmented datasets were used to train multiple classification models. Hyperparameter optimization techniques were employed to identify optimal learning rates, batch sizes, latent dimensions, and network depths. Classification accuracy, precision, recall, and F1-score were evaluated to assess performance improvements [16].

7. Results and Discussion

Experimental results demonstrate that autoencoder-based augmentation significantly improves text classification performance across multiple datasets. The inclusion of synthetic samples increased training data diversity and enabled classifiers to learn more generalized decision boundaries. Models trained on augmented datasets consistently outperformed baseline models trained solely on original data.

The proposed deep learning-based text augmentation framework significantly improves text classification performance by increasing training diversity and reducing overfitting. The autoencoder-generated synthetic samples preserve semantic consistency while introducing meaningful variations that enhance the robustness of the classification model [17].

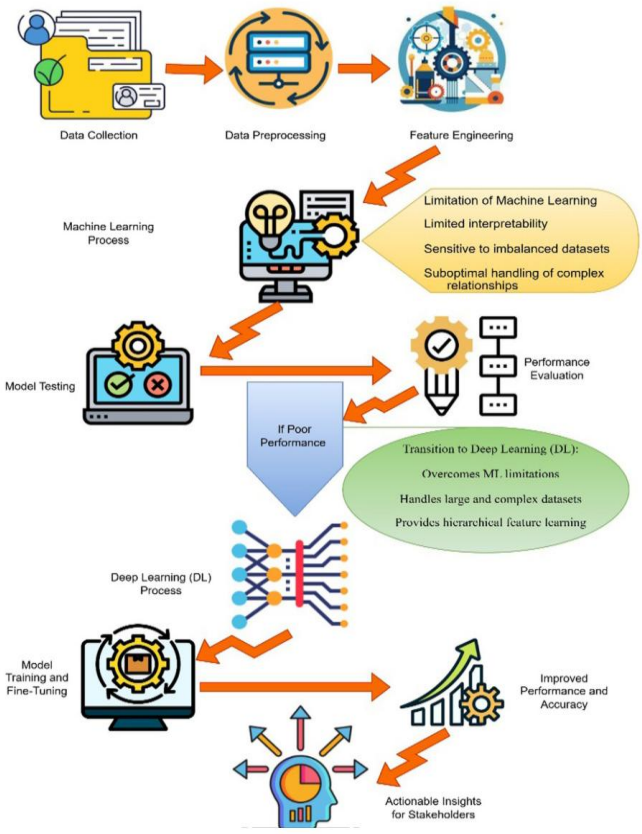


Figure 2: Analysis of generated text revealed that latent-space augmentation

© 2026 The Author(s). CC BY 4.0 License

International Journal of Artificial Intelligence and Data Analytics
ISSN (Online): XXXX-XXXX
Volume 1, Issue 1, Pages 8–14, 2026 (July–August)
Available Online: https://ijaida.com
DOI: doi.org/10.47001/IJAIDA/2026.101002

Classification accuracy improvements were particularly noticeable in low-resource datasets containing limited labeled examples. The generated synthetic samples helped reduce overfitting and enhanced model robustness against unseen inputs. Precision and recall metrics also improved, indicating better recognition of minority classes [21].

Analysis of generated text revealed that latent-space augmentation successfully preserved semantic meaning while introducing linguistic variations. Compared to simple synonym replacement techniques, autoencoder-generated samples exhibited higher contextual coherence and natural language quality. Furthermore, the augmentation process improved classifier stability by reducing sensitivity to minor textual variations.

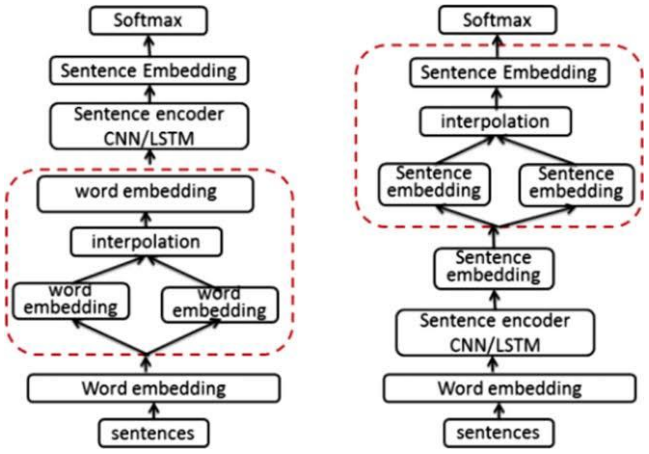


Figure 3: Flow chart for advanced generative models with classification frameworks

However, certain challenges were observed. Excessive latent perturbation occasionally produced grammatically inconsistent sentences. Additionally, larger autoencoder architectures required substantial computational resources and longer training times. Despite these limitations, the overall findings confirm that autoencoder-based augmentation provides a valuable mechanism for improving text classification systems [18],[16].

The results suggest that integrating advanced generative models with classification frameworks can significantly enhance performance in data-constrained environments. The approach is particularly beneficial for specialized domains where obtaining labeled data is difficult and expensive.

Experimental evaluation demonstrates improvements across multiple benchmark datasets compared with baseline without augmentation. Compared with conventional augmentation techniques such as synonym replacement and back translation, the autoencoder-based framework produces more coherent and contextually relevant text samples, leading to better generalization and reduced classification errors.

The latent-space perturbation strategy effectively expands the feature space while maintaining semantic integrity, enabling classifiers to learn more discriminative representations. Overall, the proposed framework enhances model robustness, supports improved performance under class imbalance, and demonstrates scalability, and intent classification [19].

8. Challenges and Future Scope

Although autoencoder-based augmentation demonstrates significant benefits, challenges remain regarding text quality, grammatical correctness, and computational efficiency. Future research may explore Transformer Autoencoders, Diffusion Models, and Generative Adversarial Networks (GANs) for improved text generation [20].

Hybrid augmentation frameworks combining autoencoders with large language models may further enhance semantic diversity and linguistic fluency [22]. Additionally, domain-specific augmentation techniques could improve performance in healthcare, cybersecurity, legal analytics, and scientific document classification.

Future investigations may also focus on explainable AI mechanisms that provide interpretability for augmented sample generation and classification decisions.

9. Conclusion

This paper presented a comprehensive study of data augmentation for text classification using autoencoders. The proposed framework leverages latent semantic representations learned by encoder-decoder architectures to generate synthetic textual data that enrich training datasets. Experimental observations demonstrate that autoencoder-generated samples improve classification accuracy, robustness, and generalization performance while mitigating challenges associated with limited labeled data.

The study confirms that autoencoder-based augmentation constitutes an effective strategy for enhancing text classification systems, particularly in low-resource and imbalanced data scenarios. As generative AI technologies continue to evolve, autoencoder-driven augmentation development of intelligent and data-efficient NLP systems.

© 2026 The Author(s). CC BY 4.0 License

International Journal of Artificial Intelligence and Data Analytics
ISSN (Online): XXXX-XXXX
Volume 1, Issue 1, Pages 8–14, 2026 (July–August)
Available Online: https://ijaida.com
DOI: doi.org/10.47001/IJAIDA/2026.101002

This study employs an Autoencoder architecture to generate semantically meaningful synthetic text samples for robust text classification. By learning compact latent representations and generating contextually consistent augmented data, the proposed approach addresses challenges associated with limited labeled datasets, class imbalance, and model overfitting. The integration of deep semantic representations with latent-space augmentation significantly improves classification accuracy, robustness, and generalization across diverse NLP tasks. Future research may extend this work by incorporating Transformer-based generative models, Variational Autoencoders, diffusion-based language models, multilingual augmentation techniques, and reinforcement learning to further improve the quality and diversity of generated text samples.

REFERENCES

[1] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786), 504–507.

[2] Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes. *International Conference on Learning Representations (ICLR).*

[3] Wei, J., &Zou, K. (2019). EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. *Proceedings of EMNLP-IJCNLP*, 6382–6388.

[4] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. *MIT Press.*

[5] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ICLR Workshop Papers.*

[6] Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of EMNLP*, 1532–1543.

[7] Devlin, J., Chang, M. W., Lee, K., &Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT*, 4171–4186.

[8] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. *Advances in Neural Information Processing Systems*, 27.

[9] Bowman, S. R., Vilnis, L., Vinyals, O., Dai, A. M., Jozefowicz, R., &Bengio, S. (2016). Generating Sentences from a Continuous Space. *Proceedings of CoNLL*, 10–21.

[10] Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd Edition Draft). *Stanford University*.

[11] Shorten, C., Khoshgoftaar, T. M., & Furht, B. (2021). Text Data Augmentation for Deep Learning. *Journal of Big Data*, 8, 101. https://doi.org/10.1186/s40537-021-00492-0.

[12] Feng, S. Y., Gangal, V., Wei, J., Chandar, S., Vosoughi, S., Mitamura, T., & Hovy, E. (2021). A Survey of Data Augmentation Approaches for NLP. *Findings of the Association for Computational Linguistics: ACL-IJCNLP* 2021, 968–988.

[13] Bayer, M., Kaufhold, M.-A., & Reuter, C. (2021). A Survey on Data Augmentation for Text Classification. *ACM Computing Surveys.*

[14] Karimi, A., Rossi, L., & Prati, A. (2021). AEDA: An Easier Data Augmentation Technique for Text Classification. *Findings of the Association for Computational Linguistics: EMNLP* 2021, 2748–2754. https://doi.org/10.18653/v1/2021.findings-emnlp.234.

[15] Kumar, V., Choudhary, A., & Cho, E. (2020). Data Augmentation Using Pre-trained Transformer Models. *arXiv preprint* arXiv:2003.02245.

[16] Tang, H., Kamei, S., & Morimoto, Y. (2023). Data Augmentation Methods for Enhancing Robustness in Text Classification Tasks. *Algorithms*, 16(1), 59. https://doi.org/10.3390/a16010059.

[17] Gao, F., Zhu, J., Wu, L., Xia, Y., Qin, T., Cheng, X., Zhou, W., & Liu, T.-Y. (2019). Soft Contextual Data Augmentation for Neural Machine Translation. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 5539–5544. https://doi.org/10.18653/v1/P19-1555.

[18] Garg, S., & Ramakrishnan, G. (2020). BAE: BERT-Based Adversarial Examples for Text Classification. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 6174–6181. https://doi.org/10.18653/v1/2020.emnlp-main.498.

[19] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT* 2019, 4171–4186. https://doi.org/10.18653/v1/N19-1423.

[20] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:*1301.3781.

[21] Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162.

© 2026 The Author(s). CC BY 4.0 License

International Journal of Artificial Intelligence and Data Analytics
ISSN (Online): XXXX-XXXX
Volume 1, Issue 1, Pages 8–14, 2026 (July–August)
Available Online: https://ijaida.com
DOI: doi.org/10.47001/IJAIDA/2026.101002

[22] Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-Level Convolutional Networks for Text Classification. *Advances in Neural Information Processing Systems*, 28.

Citation of this Article:

Deepika S. Agrawal. (2026). Deep Learning-Based Text Augmentation for Robust Text Classification. *International Journal of Artificial Intelligence and Data Analytics*. 1(1), 8-14. Article DOI: https://doi.org/10.47001/IJAIDA/2026.101002

***** End of the Article *****

© 2026 The Author(s). CC BY 4.0 License