Exploring Character-Based Stylometry Features Using Machine Learning for Intrinsic Plagiarism Detection in Urdu

Muhammad Faraz Manzoor[1,3*], Muhammad Shoaib Farooq[1], Muntazir Mehdi[2], Adnan Abid[1,4]
[1]Department of Computer Science, University of Management and Technology, Lahore, Pakistan.
[2]Department of Computer Science, Virtual University of Pakistan, Lahore, Pakistan.
[3]Department of Computer Science, Bahria University, Lahore, Pakistan.
[4]Department of Data Science, Faculty of Computing and Information Technology, University of the Punjab, Pakistan.
***Correspondence:** F2018288004@umt.edu.pk

Citation | Manzoor. M. F, Farooq. M. S, Mehdi. M, Abid. A, "Exploring Character-Based Stylometry Features Using Machine Learning for Intrinsic Plagiarism Detection in Urdu", Special Issue. pp 236-245, Oct 2024
Received | Oct 14, 2024 **Revised |** Oct 18, 2024 **Accepted |** Oct 22, 2024 **Published |** Oct 28, 2024.

Plagiarism detection in natural language processing (NLP) plays a crucial role in maintaining textual integrity across various domains, particularly for low-resource languages like Urdu. This study addresses the emerging challenge of intrinsic plagiarism detection in Urdu, an area with limited research due to the scarcity of datasets and model resources. To bridge this gap, our research investigates the use of character-based stylometric features in combination with machine learning (ML) and deep learning (DL) models specifically designed for Urdu text analysis. We conducted a series of experiments to evaluate the performance of several classifiers, including Random Forest, AdaBoost, K-Nearest Neighbor (KNN), Decision Tree, Gaussian Naive Bayes, and Long Short-Term Memory (LSTM) networks. Our results show that KNN and LSTM achieved the highest accuracy at 74%, with KNN outperforming the others in terms of F1-score (64.3%), highlighting its balanced performance across accuracy, precision, and recall. AdaBoost followed closely with an accuracy of 73% and a precision of 77.5%, although its F1-score was slightly lower at 63.6%. These findings emphasize the need for specialized approaches in NLP for Urdu, demonstrating that tailored ML and DL techniques can significantly improve intrinsic plagiarism detection in low-resource languages.

Keywords: Intrinsic; Plagiarism; Urdu; Stylometry.

Introduction

Plagiarism is a serious ethical violation that involves misrepresenting someone else's work as one's own. This unethical practice occurs when an individual knowingly uses content from another source without providing proper attribution, attempting to deceive others by falsely assuming authorship [1], [2]. As a result, many publishers and academic institutions have implemented stringent measures to prevent plagiarism, applying significant penalties to those found guilty of the offense.

Plagiarism detection can be categorized into two primary approaches: intrinsic and extrinsic. Intrinsic plagiarism detection focuses on determining whether the entire document or specific portions were written by a single author. It primarily looks for discrepancies in writing style within the text, identifying segments that differ from the overall tone or structure of the document. In contrast, extrinsic plagiarism detection involves comparing the suspect document against a corpus of known sources to identify any content—such as phrases or sentences—that appears in both the suspect text and other sources. This method typically employs algorithms that scan the web, index existing content, and then use keyword analysis to find exact matches. Recent advances in natural language processing (NLP) have enhanced the detection process, as well as facilitated plagiarism prevention. Today, a range of online and offline tools are available to detect plagiarized text. Global detection efforts are supported by popular commercial tools like Turnitin and Plagscan, which are designed to identify plagiarism after it has occurred, as they do not have the capability to prevent it [2], [3].

Various techniques are used to detect intrinsic plagiarism, such as identifying stylometry features, text segmentation, and lexical analysis. This study specifically focuses on the use of stylometry to detect intrinsic plagiarism. Stylometry refers to the analysis of linguistic style and involves examining both word-based and character-based features. These features include sentence length, punctuation, sentence structure, and other stylistic markers that contribute to an author's unique writing style. By identifying patterns in these features, stylometry can be particularly effective in detecting intrinsic plagiarism, even when the text has been paraphrased or altered. Analyzing these features helps uncover subtle similarities that might suggest plagiarism, offering insights into the text's original author.

Machine learning and deep learning techniques have proven to be successful in detecting both intrinsic and extrinsic plagiarism, particularly in the English language [5], [6], [7]. The widespread availability of English-language datasets and programming tools has made it easier to develop and deploy plagiarism detection models. However, in countries like Pakistan, where a significant amount of academic content is written in Urdu, there is a notable gap in plagiarism detection tools for this language. The lack of an intrinsic plagiarism corpus for Urdu has made it difficult to address this issue effectively. Therefore, the objective of this research is to explore and develop efficient methods for detecting intrinsic plagiarism in Urdu text using character-based stylometry features combined with machine learning and deep learning techniques. This study makes two key contributions:

First, it introduces a novel set of character-based stylometry features specifically designed for Urdu text analysis, which enhances the accuracy of intrinsic plagiarism detection in the language.

Second, it compares the performance of various machine learning, ensemble learning, and deep learning classifiers in the context of Urdu intrinsic plagiarism detection (UIPD) to identify the most effective approach for this task.

Objectives

- To assess the effectiveness of character-based stylometry features in detecting intrinsic plagiarism in Urdu text

- To compare the performance of various machine learning (ML) and deep learning (DL) classifiers, with the aim of identifying the most effective models for processing the Urdu language.

Novelty:

This study introduces a novel approach that combines character-based stylometry features with machine learning (ML) and deep learning (DL) models, specifically designed for intrinsic plagiarism detection in Urdu. By focusing on this under-explored area, the research addresses the challenges of plagiarism detection in low-resource languages and helps bridge the gap created by limited datasets and model resources.

Related Work:

A significant amount of research has been dedicated to intrinsic plagiarism detection, particularly in the English language. This section offers a brief overview of the common datasets and techniques currently used in the field, providing insight into the state of research and the methodologies employed for intrinsic plagiarism detection.

Datasets:

The Webis-CPC-11 corpus [8] contains 7,859 potential paraphrases, which were generated through Mechanical Turk crowdsourcing. This dataset includes 4,067 verified paraphrases, along with the associated source texts, as well as 3,792 non-paraphrases that were rejected. These samples were not previously available as standalone data, apart from the larger competition dataset, although they were originally used in the PAN 2010 global plagiarism detection competitions. In contrast, Potthast et al. [9] developed the Webis-PC-08 benchmark dataset, which was subsequently updated over the following years to PAN-PC-09, PAN-PC-10, and ultimately PAN-PC-11. This dataset is designed to evaluate two types of retrieval tasks in automatic plagiarism detection: (1) extrinsic plagiarism detection and (2) intrinsic plagiarism detection.

Extrinsic plagiarism detection involves a set of both suspicious documents and source documents, with the goal of identifying plagiarized portions in the suspicious documents and matching them to their corresponding sections in the source documents. In contrast, intrinsic plagiarism detection focuses solely on suspicious documents, aiming to identify all plagiarized sections, such as breaches in writing style. In this task, comparison with other documents is not permitted. Additionally, the PAN-PC-09 dataset includes documents with artificially inserted plagiarism, created using a random plagiarist program. This program constructs plagiarism cases based on various random variables, including the percentage of plagiarism across the entire corpus, the percentage of plagiarism per document, the length of individual plagiarized portions, and the complexity of each plagiarized section.

Existing Techniques:

The Webis-CPC-11 corpus [8] contains 7,859 potential paraphrases, which were generated through Mechanical Turk crowdsourcing. This dataset includes 4,067 verified paraphrases, along with the associated source texts, as well as 3,792 non-paraphrases that were rejected. These samples were not previously available as standalone data, apart from the larger competition dataset, although they were originally used in the PAN 2010 global plagiarism detection competitions. In contrast, Potthast et al. [9] developed the Webis-PC-08 benchmark dataset, which was subsequently updated over the following years to PAN-PC-09, PAN-PC-10, and ultimately PAN-PC-11. This dataset is designed to evaluate two types of retrieval tasks in automatic plagiarism detection: (1) extrinsic plagiarism detection and (2) intrinsic plagiarism detection.

Extrinsic plagiarism detection involves a set of both suspicious documents and source documents, with the goal of identifying plagiarized portions in the suspicious documents and

matching them to their corresponding sections in the source documents. In contrast, intrinsic plagiarism detection focuses solely on suspicious documents, aiming to identify all plagiarized sections, such as breaches in writing style. In this task, comparison with other documents is not permitted. Additionally, the PAN-PC-09 dataset includes documents with artificially inserted plagiarism, created using a random plagiarist program. This program constructs plagiarism cases based on various random variables, including the percentage of plagiarism across the entire corpus, the percentage of plagiarism per document, the length of individual plagiarized portions, and the complexity of each plagiarized section.

Materials:

We utilized a newly published dataset specifically designed for sentence-level intrinsic plagiarism detection in Urdu [16]. This dataset was carefully curated to ensure its relevance and effectiveness for training classification algorithms. To create a high-quality, real-world representative dataset, data was collected from a variety of reputable Urdu-language sources, including websites such as jang.com, urduessaypoint.blogspot.com, and dawnnews.tv, among others. All publications obtained from these sources were systematically compiled into a standardized .txt format. The dataset consists of 2,520 documents, evenly split into plagiarized and non-plagiarized categories. A detailed breakdown of this dataset is provided in Table 1.

Table 1. Characteristics of the dataset

Main Topic	Sub Topic	Plagiarized	Non-Plagiarized	Total
National Celebrities	Quaid e Azam	126	126	252
	Allama Iqbal	126	126	252
Annual Events	Independence Day	126	126	252
	Eid ul Fitr	126	126	252
	Importance of Forest	126	126	252
Moral Lesson	Greatness in hard work	126	126	252
	Importance of Sports	126	126	252
	Importance of Education	126	126	252
	Behavior with parents	126	126	252
Today's World	Technology	126	126	252
Total		**1260**	**1260**	**1260**

Methods:

The primary objective of this study is to classify documents as plagiarized or non-plagiarized using machine learning (ML) techniques, with a focus on character-based stylometry features. To ensure the robustness of our results, we employ six distinct classifiers: Random Forest, Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes, AdaBoost, and Long Short-Term Memory (LSTM). Additionally, we apply Principal Component Analysis (PCA) as a dimensionality reduction technique. By preserving the most significant features and reducing the overall dimensionality, PCA enhances the effectiveness and efficiency of our classifiers. To evaluate the performance of these classifiers, we use several assessment metrics, including F1-score, accuracy, precision, and recall. Furthermore, we compare our findings with those of prior research on this topic, which has employed various methodologies and languages. A detailed overview of the methodology is presented in Figure 1.

Stylometry Feature Extraction:

The quantification of writing style is achieved through the use of stylometric features [17]. Every author exhibits unique writing and typing characteristics, some of which may be used consciously, while others are employed unconsciously [18]. Since stylometric features in Urdu differ from those in English, and many stylometric features are language-dependent, we have selected features that are commonly employed by Urdu authors. These features are outlined in Table 2.

Table 2. Character based Stylometry Features

SR	Feature	SR	Feature
1	Comma count	9	Ampersands count
2	Dashes count	10	Percentage signs
3	Open parentheses count	11	Number of single quotes
4	Close parentheses count	12	Number of double quotes
5	Semicolons count	13	Colons count
6	White spaces count	14	Number of characters without spaces
7	Question marks count	15	Digit count
8	Exclamation marks count	16	No of all brackets

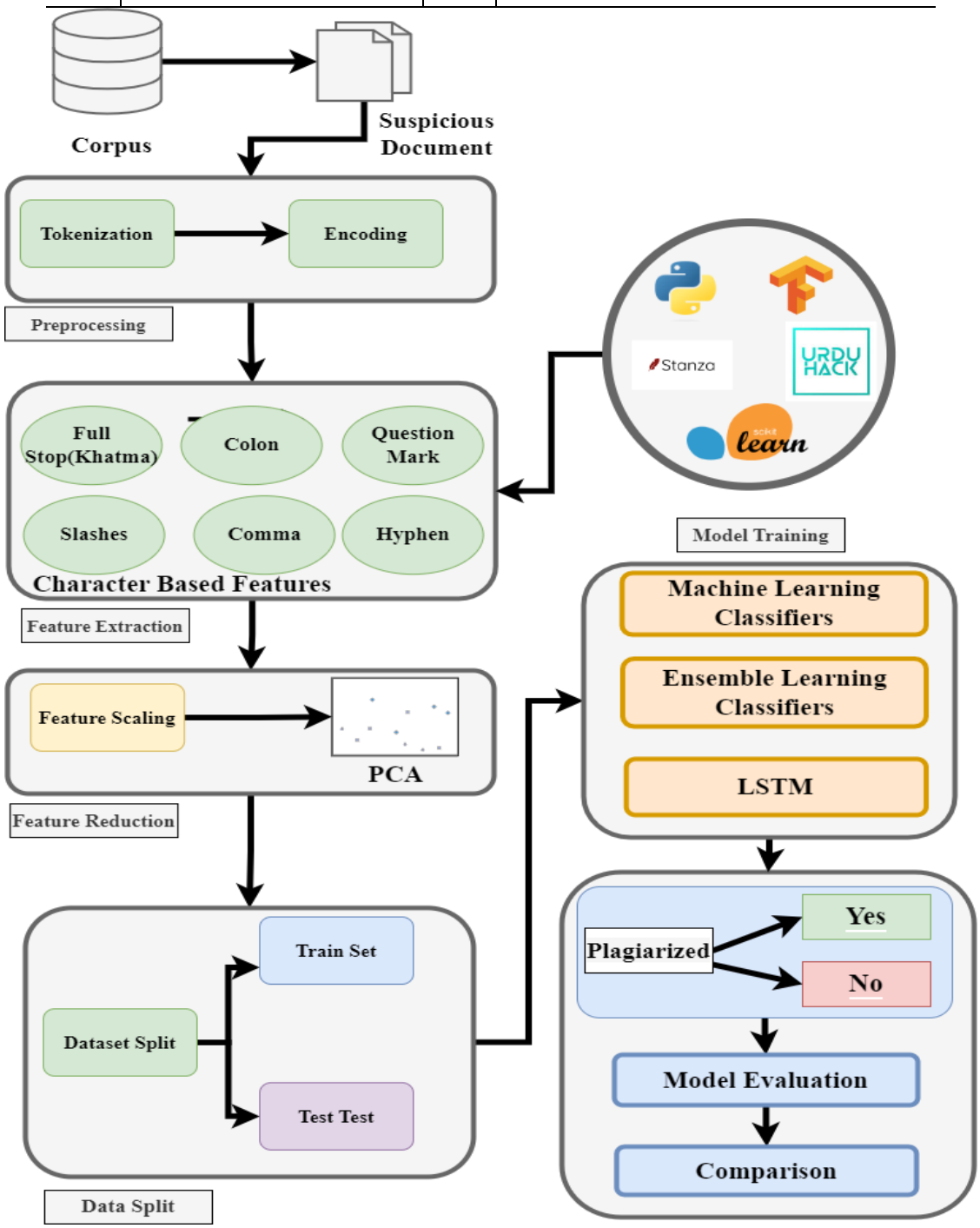

Figure 1. Methodology for intrinsic plagiarism detection in Urdu

Data Pre-Processing:

To prepare the documents for evaluation, essential preprocessing steps are performed to ensure compatibility with the classifiers. These steps typically include removing stop words and punctuation marks, as well as cleaning words containing alphanumeric characters by eliminating numeric digits. However, recognizing that such text characteristics contribute to stylometric features, and considering the resource constraints associated with Urdu text, our preprocessing approach is adapted. Specifically, we: a) break paragraphs and sentences into smaller, more manageable units for easier analysis, and b) encode the output labels of the data into numerical representations using appropriate encoding methods. Furthermore, the dataset is split into training and test sets, with 80% of the data used for training and the remaining 20% reserved for testing, allowing us to evaluate the predictive performance of the model.

Models:

Machine learning techniques have been widely used by various researchers to address the problem of intrinsic plagiarism detection, with demonstrated effectiveness in multiple languages. In this study, we employed a range of traditional machine learning, ensemble learning, and deep learning algorithms, including Random Forest [19], AdaBoost [20], Decision Tree [21], K-Nearest Neighbors (KNN) [22], Naive Bayes [23], and Long Short-Term Memory (LSTM) [24], for classifying plagiarized and non-plagiarized documents. The configuration details of these models are provided in Table 3.

Table 3. Parameters and Configurations of Models

SR#	Model	Parameters and Configurations
1	RF	n estimator=100
2	AdaBoost	n estimators=100; Learning rate=1; base estimator=DT
3	KNN	Total Neighbors=5
4	Decision Tree	Criteria =Entropy
5	Naive Bayes	--
6	LSTM	Dense Layers=2 Activation Function=Relu, Sigmoid Epochs=100 Optimizer=Adam

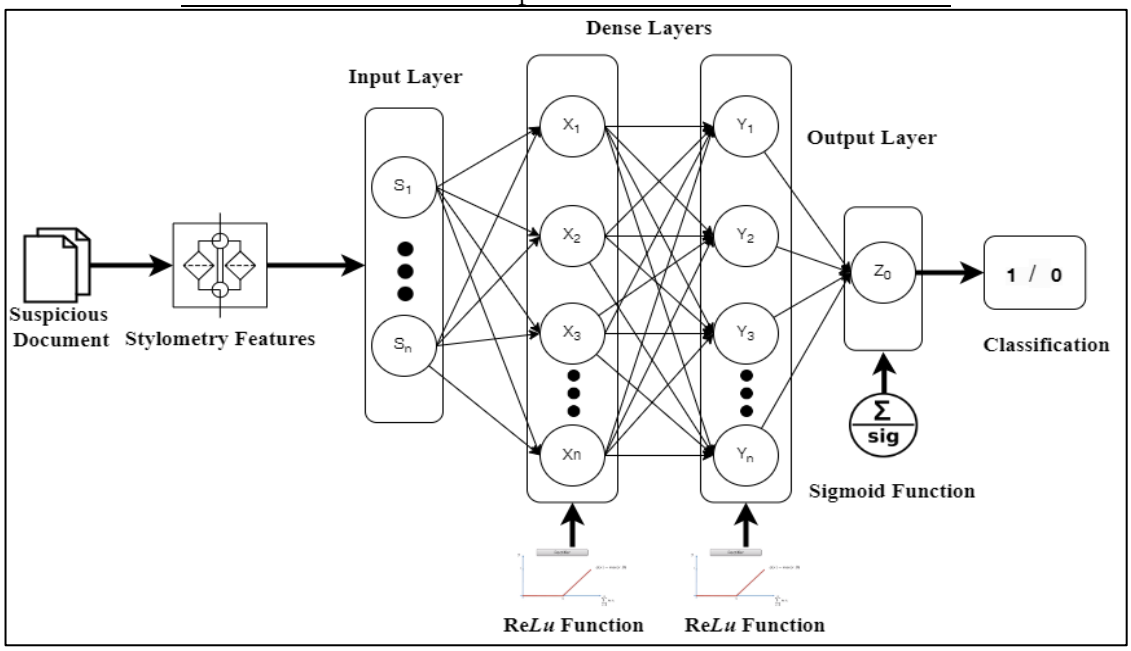

Figure 2. Propose Architecture of LSTM Model

In the LSTM architecture, traditional hidden layers are replaced with LSTM cells, which contain specialized gates to regulate the flow of information. In this research, we have enhanced the standard LSTM structure by incorporating key components, including the input gate, cell state, forget gate, and output gate, to improve the model's ability to capture long-term dependencies in the data.

The stylometric features are captured using two LSTM layers, with ReLU activation functions applied to each layer. To prevent feature co-adaptation, dropout and batch normalization are applied following the fully connected layer. A binary activation function is used to generate the final prediction for the label. The binary cross-entropy loss function is employed, and the "Adam" optimizer is selected due to its superior performance compared to other optimization techniques. The architecture of the LSTM model used in this study is illustrated in Figure 2.

Performance Measure Parameters:

We employ a comprehensive set of performance metrics to evaluate the effectiveness of our machine learning models. Specifically, we rely on accuracy, precision, recall, and the F1 score as key indicators of model performance. Accuracy measures the overall correctness of predictions by calculating the ratio of correctly predicted examples to the total number of examples. Precision assesses the model's ability to minimize false positives, reflecting the percentage of accurate positive predictions among all instances classified as positive. Recall, on the other hand, gauges the model's ability to identify valid positive samples within the entire set of actual positive cases. The F1 score serves as a balanced measure, considering the trade-off between precision and recall. By incorporating all these metrics, we ensure a thorough and robust evaluation of our models' performance and classification accuracy.

Result and Discussion:

In this study, the classification of suspicious documents was performed using character-based stylometry features. The performance of the classifiers was evaluated based on eleven word-based stylometry features extracted from the suspicious documents. As shown in Table 4, LSTM and KNN achieved the highest accuracy (74%) compared to the other classifiers we tested. Additionally, AdaBoost yielded the highest precision at 77.5%, while Decision Tree achieved the second-highest precision with 69%. It is also worth noting that Naive Bayes.

Table 4. Performance of Models on Character based Stylometry Features

Classifier	Accuracy%	Precision%	Recall%	F1-Score%
Random Forest	69	62	57	59.3
AdaBoost	73	77.5	54	63.6
KNN	74	68	61	64.3
Decision Tree	73	69	55	61.2
Naive Bayes	66	49	50	49.49
LSTM	74	55	56	55.49

Among all the classifiers tested, KNN demonstrated the best overall performance. This can be attributed to KNN being a non-parametric algorithm, which means it does not rely on any assumptions about the underlying data distribution. It is highly flexible, capable of handling large datasets without requiring knowledge of the data's quantity or form. This characteristic makes KNN particularly effective at recognizing complex patterns in high-dimensional data, such as textual data.

When compared to previous studies, our model showed superior performance in intrinsic plagiarism detection for Urdu text, as illustrated in Table 5. For instance, Stamatatos et al. [25] achieved a maximum precision of 46.07% and an F1-score of 30.86% using the Standardized Distance Function, while Kuznetsov et al. [26] obtained 44% precision and 42% F1-score with Gradient-Enhancing Regression Trees. In contrast, our KNN model

outperformed these results, achieving 74% accuracy, 68% precision, and 64.3% F1-score. Similarly, Tschuggnall et al. [27] reported a lower recall of 23% and an F1-score of 24% using Grammar Tree Comparison on scientific documents, while Alsallal et al. [28] reported a higher precision of 61.93%, though their F1-score was not disclosed. Overall, our KNN approach delivered a well-rounded performance across all metrics, highlighting the effectiveness of tailored machine learning models for intrinsic plagiarism detection, particularly in low-resource languages such as Urdu.

Table 5. Performance Comparison of This Study with Existing Techniques

Author(s)	Year	Model/Technique	Dataset	Accuracy	Precision	Recall	F1-Score
Stamatatos et al. [25]	2009	Standardized Distance Function	IPAT-CC	-	23.21%	46.07%	30.86%
Kuznetsov et al. [26]	2016	Gradient Enhancing Regression Trees	PAN 2011	-	44%	47%	42%
Tschuggnall et al. [27]	2023	Grammar Tree Comparison	Scientific documents from internet	-	-	23%	24%
Alsallal et al. [28]	2013	Various classification techniques	MED dataset	-	61.93%	80.16%	
This study	2024	KNN	Essays	74	68	61	64.3

Conclusion:

This study tackles the problem of intrinsic plagiarism detection in Urdu by systematically developing a corpus specifically designed for this task. Widely used machine learning, ensemble learning, and deep learning classifiers were applied to detect intrinsic plagiarism in the Urdu language, using character-based stylometry features. The results demonstrated that KNN outperformed other classifiers, including those based on machine learning, deep learning, and ensemble learning techniques. This research provides valuable insights into intrinsic plagiarism detection for the Urdu language and lays the groundwork for future investigations in this area.

For future work, transfer learning models could be explored to further improve the accuracy of plagiarism detection in Urdu. Additionally, the dataset could be expanded to include paragraph-level data, incorporating a broader range of sources to enhance the robustness of the model. Crowdsourcing [29] could be leveraged to generate and validate larger datasets more efficiently, ensuring timely updates and improvements to the model.

Author's Contribution: Muhammad Faraz Manzoor: Data Curation, Methodology Development, **Muhammad Shoaib Farooq:** Manuscript Preparation, Supervision, **Muntazir Mehdi**: Data Curation, **Adnan Abid**: Methodology Development

Conflict of Interest: Authors have no conflict of interest.

References:

[1] P. Samuelson, "Self-plagiarism or fair use," Commun. ACM, vol. 37, no. 8, pp. 21–25, 1994.

[2] A. Hashemi and W. Shi, "Enhancing Writing Style Change Detection using Transformer-based Models and Data Augmentation," CEUR Workshop Proc., vol. 3497, pp. 2613–2621, 2023.

[3] N. Beute, E. S. Van Aswegen, and C. Winberg, "Avoiding plagiarism in contexts of development and change," IEEE Trans. Educ., vol. 51, no. 2, pp. 201–205, 2008.

[4] P. Clough and others, "Old and new challenges in automatic plagiarism detection," Natl.

plagiarism Advis. Serv., vol. 41, pp. 391–407, 2003.
[5] M. AlSallal, R. Iqbal, V. Palade, S. Amin, and V. Chang, "An integrated approach for intrinsic plagiarism detection," Futur. Gener. Comput. Syst., vol. 96, pp. 700–712, 2019.
[6] D. Curran, "An evolutionary neural network approach to intrinsic plagiarism detection," in Artificial Intelligence and Cognitive Science: 20th Irish Conference, AICS 2009, Dublin, Ireland, August 19-21, 2009, Revised Selected Papers 20, 2010, pp. 33–40.
[7] H. R. Iqbal, R. Maqsood, A. A. Raza, and S. U. Hassan, "Urdu paraphrase detection: A novel DNN-based implementation using a semi-automatically generated corpus," Nat. Lang. Eng., pp. 1–31, 2023, doi: 10.1017/S1351324923000189.
[8] S. Burrows, M. Potthast, and B. Stein, "Paraphrase acquisition via crowdsourcing and machine learning," ACM Trans. Intell. Syst. Technol., vol. 4, no. 3, pp. 1–21, 2013.
[9] M. Potthast, A. Eiselt, L. A. Barrón Cedeño, B. Stein, and P. Rosso, "Overview of the 3rd international competition on plagiarism detection," in CEUR workshop proceedings, 2011.
[10] and A. S. Andrianna Polydouri(B), Georgios Siolas and Intelligent, "Intrinsic Plagiarism Detection with Feature-Rich Imbalanced Dataset Learning," Eng. Appl. Neural Networks, vol. 2, pp. 87–98, 2017, doi: 10.1007/978-3-319-65172-9.
[11] M. AlSallal, R. Iqbal, V. Palade, S. Amin, and V. Chang, "An integrated approach for intrinsic plagiarism detection," Futur. Gener. Comput. Syst., vol. 96, pp. 700–712, 2019, doi: 10.1016/j.future.2017.11.023.
[12] C. Zuo, Y. Zhao, and R. Banerjee, "Style change detection with feed-forward neural networks notebook for PAN at CLEF 2019," CEUR Workshop Proc., vol. 2380, no. September, pp. 9–12, 2019.
[13] J. A. Khan, "Style breach detection: An unsupervised detection model: Notebook for PAN at CLEF 2017," CEUR Workshop Proc., vol. 1866, 2017.
[14] A. Saini, M. R. Sri, and M. Thakur, "Intrinsic plagiarism detection system using stylometric features and DBSCAN," Proc. - IEEE 2021 Int. Conf. Comput. Commun. Intell. Syst. ICCCIS 2021, pp. 13–18, 2021, doi: 10.1109/ICCCIS51004.2021.9397187.
[15] J. Brooke and G. Hirst, "Paragraph Clustering for Intrinsic Plagiarism Detection using a Stylistic Vector Space Model with Extrinsic Features.," CLEF (Online Work. Notes/Labs/Workshop), pp. 1–9, 2012, [Online]. Available: http://ceur-ws.org/Vol-1178/CLEF2012wn-PAN-BrookeEt2012.pdf
[16] F. Manzoor, M. S. Farooq, A. Abid, and A. Alvi, "Language Resources for Intrinsic Plagiarism Detection in Urdu Language," Mendeley Data, 2023, doi: 10.17632/8fknny5s5p.2.
[17] K. Lagutina et al., "A survey on stylometric text features," in 2019 25th Conference of Open Innovations Association (FRUCT), 2019, pp. 184–195.
[18] S. Adamović et al., "An efficient novel approach for iris recognition based on stylometric features and machine learning techniques," Futur. Gener. Comput. Syst., vol. 107, pp. 144–157, 2020.
[19] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," Expert Syst. Appl., vol. 134, pp. 93–101, 2019.
[20] A. Vezhnevets and V. Vezhnevets, "Modest AdaBoost-teaching AdaBoost to generalize better," in Graphicon, 2005, pp. 987–997.
[21] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," J. Chemom. A J. Chemom. Soc., vol. 18, no. 6, pp. 275–285, 2004, doi: https://doi.org/10.1002/cem.873.
[22] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in On The Move to Meaningful Internet Systems 2003: CoopIS, DOA,

and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings, 2003, pp. 986–996.

[23] A. H. Jahromi and M. Taheri, "A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features," in 2017 Artificial intelligence and signal processing conference (AISP), 2017, pp. 209–212.

[24] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," Neural Comput., vol. 31, no. 7, pp. 1235–1270, 2019.

[25] E. Stamatatos, "Intrinsic plagiarism detection using character n-gram profiles," threshold, vol. 2, no. 1,500, 2009.

[26] M. P. Kuznetsov, A. Motrenko, R. Kuznetsova, and V. V Strijov, "Methods for Intrinsic Plagiarism Detection and Author Diarization.," in CLEF (Working notes), 2016, pp. 912–919.

[27] M. Tschuggnall and G. Specht, "Detecting plagiarism in text documents through grammar-analysis of authors," Datenbanksysteme für Business, Technol. und Web 2028, 2013.

[28] M. Alsallal, R. Iqbal, S. Amin, and A. James, "Intrinsic plagiarism detection using latent semantic indexing and stylometry," in 2013 Sixth International Conference on Developments in eSystems Engineering, 2013, pp. 145–150.

[29] H. S. Alenezi and M. H. Faisal, "Utilizing crowdsourcing and machine learning in education: Literature review," Educ. Inf. Technol., vol. 25, no. 4, pp. 2971–2986, 2020.