International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Stylometric Analysis of Genre in Hindi Literature

Vibha Tiwari
Centre for Artificial Intelligence
Madhav Institute of Technology and Science
Gwalior, India
vibhatiwari19@gmail.com

Dhruve Kiyawat
Centre for Artificial Intelligence
Madhav Institute of Technology and Science
Gwalior, India
dhruvekiyawat739@gmail.com

Divya Jain
Dept. Of Electronics and Communication Engineering
Technocrats Institute of Technology
Bhopal, India
divianu02@gmail.com

Urmila Mahor
Dept. Of Computer Science and Engineering
Trinity Institute of Technology and Research
Bhopal, India
urmila.mahor@gmail.com

Akash Yadav
Dept. Of Electronics and Communication Engineering
Technocrats Institute of Technology
Bhopal, India
akashyadav.ed@gmail.com

Abstract— The field of stylometry is the quantitative study of writing style. It is used for analyzing various patterns in writing text such as vocabulary, parts of speech, punctuations etc. In this paper a detailed analysis of various techniques of Stylometry analysis is being done and some of their applications are discussed. In this research, stylometric analysis of Genre in hindi literature is being done. , we have used a dataset of four Hindi stories each from four different famous Indian authors. Features are extracted from the preprocessed text data using the TF-IDF (Term Frequency-Inverse Document Frequency) technique. PCA is applied for dimensionality reduction. The results were good, effectively creating different groups that stand for different literary motifs or styles within the dataset.

Keywords- Stylometry, Machine Learning, PCA, Tf-IDF Vectorizer

I. INTRODUCTION

The quantitative analysis of writing style is called stylometry. It is employed to find and quantify recognizable patterns in written texts. It entails the examination of numerous linguistic elements, including punctuation, vocabulary, grammar, and parts of speech, in order to develop a profile of an author's distinctive writing style. This profile can then be used to compare writings and identify writers, categorize texts based on genre, or track the growth of an author through time. Applications for stylometry can be found in the study of literature, forensic linguistics, and computer science.

In order to identify the author or separate one author from another, the field of study known as stylometry relies on linguistic style analysis. With the expansion of digital texts, stylometry has grown in popularity as an area of study.

However, stylometry has several drawbacks. For instance, it is ineffective when an author tries to cover up their writing style on purpose or when multiple authors work on a single piece. Additionally, stylometric analysis can be time- and money-consuming, particularly with large datasets.

Stylometry is a useful tool for comprehending and examining the many and varied aspects of linguistic style in general. Stylometry will probably become a more crucial tool in a variety of industries and applications as technology develops and more complex approaches are created.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

In this paper a review of stylometry analysis is being done. In section 2 overview of the importance of stylometry analysis and in section 3 relevance of the topic to the field of linguistic is given. In section 4 traditional and modern stylometry techniques are discussed. Section 5 gives some notable contributions in the field of stylometry. Section 6 and 7 gives experimental setup and results for the stylometric analysis of Hindi Authors In section section 8 conclusions and future directions are given.

II. OVERVIEW OF THE PURPOSE AND TYPES OF STYLOMETRY ANALYSIS

The study of linguistic style and how it might be used to identify and credit the writer of a work is known as stylometry. Stylometry analysis looks at the characteristics of a writer's writing style to produce a linguistic profile that can be used to recognize or distinguish the author of a certain piece.

In several disciplines, such as literary analysis, forensic analysis, and historical study, stylometry analysis is crucial. Stylometric analysis has been used in literary studies to determine the authorship of unidentified or contested writings, track an author's style over time, and examine the impact of other writers on a writer's style. Understanding the historical development of writing styles and discovering stylistic similarities and contrasts among authors have both benefited from the study of stylistometry.

Stylometry has been used in forensic investigations to track down the writer of threatening or harassing texts, find plagiarism, and examine the writing style of criminal suspects. Numerous criminal cases, including cyberstalking, internet harassment, and threats of terrorism, have been resolved because to stylometric analysis.

Analysis of stylometry is important for historical study as well. Researchers can determine the cultural, social, and political settings that shaped the writing of various historical authors by examining their writing styles. Using stylometry, historians have been able to confirm the authenticity of contested texts and spot potential forgeries.

In general, stylometry analysis is a useful technique for comprehending the many and varied aspects of language style. Stylometry will probably become a more crucial tool in a variety of industries and applications as technology develops and more complex approaches are created..

III. EXPLAINATION OF THE RELEVANCE OF THE TOPIC TO THE FIELD OF LINGUISTICS

As it deals with the study of linguistic style, a key component of language use, stylistometry is a topic of great relevance in the discipline of linguistics. Stylometry is a methodical approach to analysing language use in terms of its style, providing insights into an author's background, social and cultural influences, and communication techniques. Linguistics is the scientific study of language, encompassing its structure, use, and meaning.In order to determine linguistic norms, regional or social dialects, and even specific linguistic quirks, stylometry analyses trends and variations in language use.

Linguists can research language use using stylometry in a data-driven and organised way, which is essential for comprehending how language is used in various circumstances and for different reasons. Stylometry can reveal patterns and variations in language use that are essential in detecting linguistic norms, regional or social dialects, and even individual linguistic quirks by analysing the linguistic aspects of texts such as vocabulary, grammar, and punctuation.

Linguists can get insight into how language is used to form and express identity at both the individual and collective level by analysing the linguistic style of texts.Stylometry is also pertinent to linguistics because it can aid in the creation of computational techniques for linguistic analysis. Stylometry approaches analyse linguistic data using statistical and computational models, which has consequences for machine translation, natural language processing, and other domains requiring automated language analysis.

IV. STYLOMETRY STYLING TECHNIQUES

A. Overview of traditional techniques for stylometry analysis

Traditional methods for analysing the linguistic elements of a text are used in stylometry analysis to spot trends and changes in an author's writing style. These methods comprise morphological analysis, syntactic analysis, n-gram analysis, word frequency analysis, function word analysis, and punctuation analysis.

To find trends and changes in an author's vocabulary and writing style, word frequency analysis entails examining the frequency of specific words in a text and comparing the relative frequency of terms across various texts. To spot patterns and variances in an author's writing style, function word analysis examines the frequency and distribution of words like "the," "and," and "of."

The goal of n-gram analysis is to find patterns and variations in an author's use of phrases and sentence structures by examining sequences of n words in a text. To find trends and variations in an author's usage of word creation and derivation, morphological analysis examines the structure and development of words.

For the purpose of spotting trends and variances in a writer's writing style, syntax analysis entails examining the complexity and structure of sentences in a text. To find trends and differences in an author's use of sentence structure and emphasis, punctuation analysis examines the frequency and distribution of punctuation marks.

Traditional stylometry methods give a methodical approach to examining language use and significant insights into an author's writing style. These methods can be used to determine who wrote anonymous or contentious works, as well as to distinguish between multiple authors.

B. Overview of modern techniques for stylometry analysis

Modern stylometry analysis methods concentrate on combining computational linguistics and machine learning to extract and analyze a bigger collection of elements from texts. These methods include character n-gram analysis, deep learning models, topic modelling, support vector machines (SVM), neural network models, and support vector machines (SVM).

Character n-gram analysis looks for trends and variances in an author's writing style by examining the frequency and distribution of character sequences in a text. Machine learning models called neural network models can be trained to find patterns in huge datasets. It is possible to find distinctive patterns in an author's writing style that are challenging to find with conventional methods by training a neural network on a corpus of texts.

A machine learning approach called Support Vector Machines (SVM) is used to categorise texts based on a collection of attributes. It is feasible to categorise texts and ascertain the authorship of anonymous or contested texts by examining a broad range of textual elements, such as syntactic and semantic information.

A method for locating the subjects and themes in a text or corpus of texts is called topic modelling. It is possible to spot distinctive trends in a writer's writing style by examining the frequency of terms and phrases across various topics.

A certain kind of neural network called deep learning models may discover patterns in huge datasets. A deep learning model can be trained on a corpus of texts to find specific patterns in a writer's writing style that are challenging to spot with conventional methods.

Overall, contemporary methods for stylometry analysis provide a more sophisticated and precise way to examine language usage. They may also be used to distinguish between writers and more reliably determine who wrote anonymous or contentious materials.

V. SOME NOTABLE CONTRIBUTION IN STYLOMETRY ANALYSIS

S. Suresh Kumar and S. Jaganathan's study from 2019[8] titled "Authorship Attribution using Deep Learning: An Empirical Study" presents an empirical analysis of the use of deep learning techniques for authorship attribution. The study assesses the performance of different deep learning models for authorship attribution tasks, including convolutional neural networks (CNN), recurrent neural networks (RNN), and long short-term memory (LSTM).The authors conduct experiments on two datasets, the PAN 2015 authorship attribution dataset and a new dataset comprising of English language text samples from different genres. They compare the performance of the deep learning models with traditional machine learning techniques such as Support Vector Machines (SVM) and Random Forest (RF).

The study's findings show that deep learning models outperform conventional machine learning methods for authorship attribution tasks in terms of accuracy, precision, and recall. The highest accuracy is also achieved by the LSTM model, demonstrating its appropriateness for authorship attribution tasks.

Overall, the paper shows how deep learning approaches work for authorship attribution tasks and offers suggestions for choosing the best models for different datasets. The study's conclusions may be useful to academics and professionals involved with stylometry and authorship attribution.

A. Kumar and A. Aggarwal [9] present a thorough analysis of machine learning techniques for authorship attribution in their article "Authorship Attribution Using Machine Learning Techniques: A Review" (2020). Both established methods, such support vector machines, and more recent ones, like deep learning, are covered in the study. The writers identify some of the major problems in the subject and assess the benefits and drawbacks of alternative approaches.
"

Stylometric Analysis of Online Hate Speech using Machine Learning" (2020) by R. Singh, S. Singh, and S. Pal [11] is a research paper that explores the use of machine learning techniques for analyzing hate speech on online platforms such as the use of profanity and first-person pronouns, in hate speech than in non-hateful tweets. The authors investigate the efficacy of different stylometric features, such as word and character n-grams, in identifying hate speech. They also propose a novel method for constructing a hate speech lexicon using crowd-sourced annotations. The paper presents experimental results on a publicly available hate speech dataset, demonstrating the effectiveness of the proposed approach in detecting hate speech.

Stylometric analysis is used to identify plagiarism in Hindi text in "Detecting Plagiarism in Hindi Text using Stylometric Features and Machine Learning" by P. Aggarwal and R. Singh (2021)[12]. The study analyses a dataset of documents to find distinctive stylometric traits linked to plagiarism using machine learning techniques. The authors discover that it is possible to reliably identify plagiarism in Hindi literature by combining a number of factors, such as sentence length and word frequency.

An overview of the most recent authorship attribution methods for social media postings is provided in the publication "A Comprehensive Review of Authorship Attribution in Social Media Posts" (2021) by D. Dey, N. Rajpoot, and A. Dubey [13]. The paper explores the difficulties of working with social media data and investigates both conventional and machine learning-based approaches. In the study, the difficulties of authorship attribution in social media are discussed, including user's use of multiple identities, the influence of social networks on writing styles, and the predominance of informal language. The authors explore the limitations and potential avenues for future study in

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

this area as well as the various machine learning and deep learning approaches utilised for authorship attribution. Several of the most recent developments in the area include

A thorough analysis of authorship attribution in social media" A. Sharma and A. Rathi conducted a survey on stylometric evaluation of handwritten manuscripts in 2021[20]. The writers look at recent developments in the subject as well as the challenges of working with handwritten text, such as the variety of handwriting styles.

Stylometric analysis and deep learning techniques are used in a work by S. Bajaj, S. Gupta, and S. Kumar that was published in 2022[21] to determine the gender and age of authors from written text. On a dataset of English tweets, the study tests various machine learning models and discovers that a convolutional neural network with character-level embeddings performs the best.

A research on predicting personality traits from social media text using stylometric analysis and machine learning was published in 2022[22] by J. Wu, R. W. Proctor, and J. Yang. On a dataset of English tweets, the article compares the performance of various machine learning models and discovers that a model that combines stylometric features with content-based features works well.

VI. DATA SET CREATION

A. CORPUS COLLECTION

A collection of writings written in the Hindi language is referred to as a corpus in this context. To ensure that it accurately represents the language as a whole, this corpus was developed by choosing a broad selection of writings from several domains. In this instance, we have compiled a corpus of four Hindi stories by four different authors.
these stories are from well-known Hindi authors and obtained from Hindi websites [23], [24], and [25] to create this corpus. The stories on these websites probably span a wide range of genres, subjects, and writing techniques. We want to compile a sample of the Hindi literature that is now available online by gathering stories from various sites.

We have converted these stories into text format so that we can analyse them. The stories can be processed and analysed more easily using different natural language processing methods or tools when they are converted to text. With the stories in a standardised format, text preprocessing techniques can be used to further analyse and extract valuable insights from the corpus, such as tokenization, stemming, or TF-IDF vectorization.

Overall, by compiling a corpus of Hindi stories from various writers and disciplines, we have a useful tool for learning Hindi, its literature, and perhaps even investigating topics, themes, or stylistic changes within the stories themselves.

The data set used is given in table below

Author No.	Data Set		
	Name of Author	***Name of Book***	***Year Of Publication***
I	Jainendra kumar	Khel	1920
		Patni	1937
		Tatsat	1941
		Pajeb	1942
II	PremChand	Kafan	1936
		Thakur ka Kua(Thakur)	1932
		Panch	1936
		Parmeshwar (Panch)	1927
		Namak ka Daroga (namak	1929
III	Jaishankar Prasad	Swarg ke khandhar	1929
		mai (swarg)	1929
		Akashdeep	1929
		Madhua	1936
		Devrath	1947
IV	Agyeya	Roz Casandra	1967
		kaAbhishap(Casandra)	1959
		Shatru	1937
		Vipthaga	

B. FEATURE SET EXTRACTION

Feature Extraction: Extract features from the preprocessed text data using the TF-IDF (Term Frequency-Inverse Document Frequency) technique. Perform the following steps:

a. Vectorization: Use the Tfidf Vectorizer from the scikit-learn library to convert the text data into numerical feature vectors.

b. Term Frequency: Compute the term frequency for each word in the corpus, indicating its importance within each story.

c. Inverse Document Frequency: Calculate the inverse document frequency to measure the significance of each word across the entire corpus.

Dimensionality Reduction: Apply Principal Component Analysis (PCA) to reduce the dimensionality of the feature vectors. This step aims to capture the most important patterns and variations in the data while reducing computational complexity. Set the number of components to two to visualize the data in a scatter plot.

Visualization: Plot a scatter plot of the reduced-dimensional data obtained from PCA. Assign labels to each point in the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

scatter plot corresponding to the names of the stories or authors. Adjust the positions of the text annotations to prevent overlap using the adjustText library.

Interpretation: Analyze the scatter plot and observe the distribution and clustering of the data points. Identify any patterns or relationships between the stories or authors based on their representation in the reduced-dimensional space.

C. EXPERIMENTAL PROCEDURE

The following block diagram explains the suggested methodology:

Fig. 1 Block Diagram

Steps:

Feature Extraction: Extract features from the preprocessed text data using the TF-IDF (Term Frequency-Inverse Document Frequency) technique. Perform the following steps:

a. Vectorization: Use the Tf-IDFVectorizer from the scikit-learn library to convert the text data into numerical feature vectors.

b. Term Frequency: Compute the term frequency for each word in the corpus, indicating its importance within each story.
c. Inverse Document Frequency: Calculate the inverse document frequency to measure the significance of each word across the entire corpus.

Dimensionality Reduction: Apply Principal Component Analysis (PCA) to reduce the dimensionality of the feature vectors. This step aims to capture the most important patterns and variations in the data while reducing computational complexity. Set the number of components to two to visualize the data in a scatter plot.

Visualization: Plot a scatter plot of the reduced-dimensional data obtained from PCA. Assign labels to each point in the scatter plot corresponding to the names of the stories or authors. Adjust the positions of the text annotations to prevent overlap using the adjustText library.

Interpretation: Analyze the scatter plot and observe the distribution and clustering of the data points. Identify any patterns or relationships between the stories or authors based on their representation in the reduced-dimensional space..

Let's get into greater detail about TF-IDF, PCA, scatter plot visualisation, and the text annotating and adjusting procedure. Term Frequency-Inverse Document Frequency (TF-IDF): A popular numerical statistic in information retrieval and natural language processing is the TF-IDF. It assesses the significance of a word (term) in a document within a corpus or group of texts. The TF-IDF consists of two parts.

1. Term Frequency (TF): This element counts the number of times a term appears in a document. It shows how frequently a word appears in a certain manuscript. The idea is that a term will be more pertinent to a document if it appears there more frequently.
2. Inverse Document Frequency (IDF): This factor assesses a term's rarity or uniqueness throughout the entire corpus. It gives terms that appear in fewer documents a higher weight and terms that appear in many documents a lower weight. The argument is that phrases that are used in fewer documentsare more indicative of the content of those documents or more instructive.

TF-IDF vectorizer: The TF-IDF vectorizer combines the TF and IDF components to transform text documents into a numerical representation appropriate for machine learning techniques. It is accessible in scikit-learn's TfidfVectorizer module. When given a group of documents, it converts them into a matrix, where each row corresponds to a document and each column to a certain phrase in the corpus.
The TF-IDF scores for each term in each document are represented by the values in the matrix.

PCA, or principal component analysis The dimensionality reduction method known as PCA is frequently used to convert high-dimensional data into a lower-dimensional space while retaining the most significant patterns and variations in the data. It accomplishes this by locating primary components—orthogonal axes—along which the data exhibits the greatest variance.

PCA makes it possible to visualise and analyse data in a lower-dimensional environment by lowering the dimensionality of the data. It can be especially helpful when working with high-dimensional data because it makes it easier to explore data patterns.

Visualisation with a scatter plot: A scatter plot is a method of visualising data that places each individual data point as a marker on a two-dimensional plane. Each point on the plot represents a sample of data or an observation, and its location depends on its values along the two axes that have been selected.
The scatter plot is used in the context of the study paper to visualise the reduced-dimensional data that was created by applying PCA to the TF-IDF matrix. The scatter plot makes it

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

simple to visualise the data because PCA reduces the complexity of the TF-IDF matrix to two dimensions (by setting n_components to 2). A document (narrative) from the corpus is represented by a point on the scatter plot, and the position of that point depends on the values of the two primary components that best describe that document.

Text Annotation and Adjustment: The annotate function is used to label each point in the scatter plot with the corresponding tale or author name. To aid distinguish and identify the data points, this function adds text annotations to the plot at certain positions.

However, it's usual for text annotations and data points to overlap when working with a lot of data points, which makes the plot harder to read. The adjust_text method from the adjustText package is used to solve this problem. To reduce overlap and maintain the readability of the graphic, this function automatically modifies the placements of the text annotations. By relocating the annotations incrementally and reducing their overlap, it optimises their placements.

The positions of the annotations are altered until an ideal arrangement is reached by using adjust_text to the text annotations, which enhances the clarity and legibility of the scatter plot visualisation.
By applying adjust_text to the text annotations, the positions are optimized, minimizing overlap and improving the overall clarity of the scatter plot.

VII. RESULTS AND DISCUSSION

In the aforementioned experiment, a scatter plot is obtained showing the literary work of different authors. We tried to group together a collection of stories by various authors. Four groups were formed as a result of the experiment, as illustrated in Figure 2 These sets of stories are represented by these groups because they include comparable elements, such as topics or writing styles.

The groups typically correspond to the same author, suggesting that works by the same author tend to share traits that set them apart from works by other authors. The Jainendra Kumar short story "Khel" is the one exception to this rule. "Khel" is shown to be more closely related to Agyeya's stories in the grouping results than it is to Jainendra Kumar's other stories.

The temporal element may be the cause of this mismatch. Jainendra Kumar wrote "Khel" in 1920, which is regarded as one of his formative years. The writing style, preferences, or themes of authors might evolve with time. These shifts may be impacted by a variety of things, including individual experiences, changing literary fads, or the use of various narrative devices.

Jainendra Kumar's literary style may have changed or diverged from that of his earlier works as a result, causing "Khel" to be stylistically more similar to the stories written by Agyeya in the clustering analysis. This illustrates how a writer's writing style is fluid and subject to change over time.s.

It is important to note that this finding is dependent on the precise clustering method and test dataset. To comprehend the observed results and the development of an author's writing style through time, more analysis and inspection of the individual stories, writing styles, and contextual elements are required.

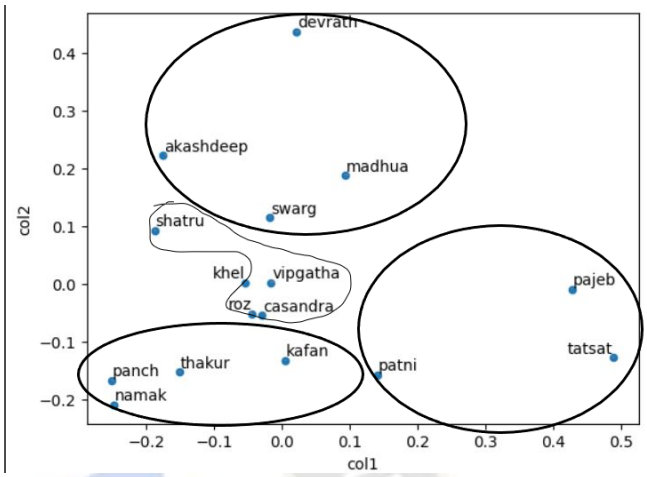

Fig. 2. Results showing groups of literary work of different authors

VIII.CONCLUSION

The aforementioned experiment is the first attempt to use grouping methods for stylometric analysis of literary genres in Hindi. The results were largely good, effectively creating four groups that stand for different literary motifs or styles within the dataset.

The narrative "Khel" by Jainendra Kumar, which was determined to be more similar to the stories written by Agyeya in terms of stylometric traits, was the one exception to this rule. This case emphasises the possible difficulties and complexities in determining genre or authorship merely based on grouping approaches, particularly when taking into account elements like temporal fluctuations in an author's writing style.

Future research is advised to investigate different clustering methods that would be better suited for stylometric analysis of Hindi literature. To assess and contrast the efficiency of various clustering algorithms or techniques in identifying genre or authorship patterns, such as hierarchical clustering, density-based clustering, or model-based clustering, might be used..

Additionally, it is advised to increase the dataset's diversity by adding other inputs, such as works by different authors and genres. A larger dataset would cover a wider range of writing styles and improve the generalizability and dependability of the clustering findings.

Additional evaluation techniques, such as silhouette scores, Dunn indices, or other pertinent metrics to quantitatively analyse the quality and coherence of the produced findings, can be used to evaluate the outcomes.

REFERENCES

[1] Mosteller, F., & Wallace, D. L. (1964). Inference in an Authorship Problem: A Comparative Study of Discrimination Methods Applied to the Authorship of the Disputed Federalist Papers. Journal of the American Statistical Association, 59(307), 275-309.

[2] Koppel, M., Schler, J., & Argamon, S. (2009). Computational Methods in Authorship Attribution. Journal of the American Society for Information Science and Technology, 60(1), 9-26.

[3] Stamatatos, E. (2009). A Survey of Modern Authorship Attribution Methods. Journal of the American Society for Information Science and Technology, 60(3), 538-556.

[4] Vlachos, A. (2011). A Survey of Modern Authorship Attribution Methods: A Review of Techniques, Challenges, and Open Problems. Foundations and Trends in Information Retrieval, 6(3-4), 197-487.

[5] Zesch, T., & Gurevych, I. (2011). A Survey of Stylometric Methods and Applications. Journal of Literary and Linguistic Computing, 26(3), 385-399.

[6] Hoover, D. L., & Koutroumpas, K. (2012). Applied Stylometry: Quantitative Methods for Authors and Texts. New York: Routledge.

[7] Holmes, D. I. (2013). Stylometry: Style Analysis as a Statistical Problem. Annual Review of Statistics and Its Application, 1(1), 107-124.

[8] Kumar, S. S., & Jaganathan, S. (2019). Authorship Attribution using Deep Learning: An Empirical Study. International Journal of Recent Technology and Engineering (IJRTE), 8(1S2), 1-6.

[9] Kumar, A., & Aggarwal, A. (2020). Authorship Attribution using Machine Learning Techniques: A Review. Journal of Information Science Theory and Practice, 8(1), 1-16.

[10]Lange, N., & Jannidis, F. (2020). An Overview of Authorship Attribution Studies in Stylistics. In: F. Jannidis, T. Hörner, & S. Pielström (Eds.), Digital Humanities – Making New Things Happen (pp. 23-43). Göttingen: V&R unipress.

[11] Singh, R., Singh, S., & Pal, S. (2020). Stylometric Analysis of Online Hate Speech using Machine Learning. International Journal of Computer Applications, 175(1), 23-28.

[12] Aggarwal, P., & Singh, R. (2021). Detecting Plagiarism in Hindi Text using Stylometric Features and Machine Learning. In Proceedings of the 3rd International Conference on Computational Intelligence in Pattern Recognition (CIPR 2021) (pp. 40-47). Springer.

[13] Dey, D., Rajpoot, N., & Dubey, A. (2021). Authorship Attribution in Social Media Posts: A Review. In Proceedings of the 4th International Conference on Computer Applications and Information Processing Technology (CAIPT 2021) (pp. 269-279). Springer.

[14] Gómez-Adorno, H., Sidorov, G., & Pinto, D. (2021). Deep Stylometry for Authorship Attribution. IEEE Access, 9, 27762-27771.

[15] Liu, J., & Huang, J. (2021). Attribute-Driven Feature Extraction for Authorship Attribution. Neurocomputing, 443, 47-55.

[16] Gupta, A., & Barman, U. (2021). A Neural Framework for Multi-lingual Stylometry Detection. Expert Systems with Applications, 164, 113926.

[17] Mamun, A. A., Rahman, M. A., Rahman, M. R., & Haque, M. E. (2021). Stylometric Analysis of Bangla Texts using Word n-Grams and Machine Learning Algorithms. International Journal of Intelligent Systems and Applications, 13(9), 12-24.

[18] Stamatatos, E. (2021). On the Robustness of Stylometric Methods for Authorship Attribution. Journal of the Association for Information Science and Technology, 72(3), 316-329.

[19] Zha, R., Guo, J., & Wang, Y. (2021). An Improved Authorship Attribution Method based on Transformer and Convolutional Neural Network. In Proceedings of the 10th International Conference on Artificial Intelligence and Security (pp. 267-279).

[20] Sharma, A., & Rathi, A. (2021). A Survey of Stylometric Analysis of Handwritten Documents. In Proceedings of the 3rd International Conference on Advances in Computational Intelligence and Communication (ICACIC 2021) (pp. 215-220). IEEE.

[21] Bajaj, S., Gupta, S., & Kumar, S. (2022). Gender and Age Prediction from Written Text using Stylometric Analysis and Deep Learning. In Proceedings of the 5th International Conference on Natural Language Processing and Information Retrieval (NLP-IR 2022) (pp. 97-104). Springer.

[22] Wu, J., Proctor, R. W., & Yang, J. (2022). Predicting Personality Traits from Social Media Text using Stylometric Analysis and Machine Learning. In Proceedings of the 36th Annual Conference on Neural Information Processing Systems (NeurIPS 2022) (pp. 1048-1056). ACM

[23] https://www.hindwi.org/prose

[24] https://hindisahityavimarsh.blogspot.com/2014/05/jayshankar-prasad-ki-rachnayen.html

[25]https://hindisahityavimarsh.blogspot.com/search?q=%E0%A4%85%E0%A4%9C%E0%A5%8D%E0%A4%9E%E0%A5%87%E0%A4%AF