V. Lakshmi Chaitanya,2023, 11:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Plagiarism Detection Process Using AI

V. Lakshmi Chaitanya, S. Nafisa Afreen, K. Veena, P. Gayathri, S. Pavitra , M. Aparna
Department of Computer Science & Engineering
Santhiram Engineering College, India

Abstract- Plagiarism relates to the act of taking information or ideas of someone else and demands it as your own. Basically, it reproduces the existing information in modified format. In every field of education, it becomes a serious issue. Various techniques and tools are derived these days to detect plagiarism. Various types of plagiarism are there like text matching, copy paste, grammar based method etc. This project proposes a new method implemented in a program. Here we put the concept of a machine learning techniques i.e. Longest Common Subsequence (LCS) and Five Modulus Method (FMM). This project helps us to identify whether text or image is plagiarized or not.

Keywords- FMM, LCS etc.

I. INTRODUCTION

1.Artificial Intelligence

Artificial Intelligenceis a branch of Computer Science and Engineering which is used to make the machine to think like human and act like human. These machines consists of sensors and actuators. Sensors helps to sense the surrounding information and the actuators act according to the information sensed by the sensors, i.e., the environment of the system is observed by intelligent agents through sensors and actuators are components through which energy is converted into motion.

They perform the role of controlling and moving a system.AI can automate workflows and processes or work independently and autonomously from a human team, it can eliminate manual errors in data processing, analytics, assembly in manufacturing, and other tasks through automation and algorithms that follow the same processes every single time, it can be used to perform repetitive tasks, freeing human capital to work on higher impact problems, AI can process more information more quickly than a human. The various applications of AI are speech recognition, image recognition, translation, productive modelling, data analytics and cyber security.

2. Longest Common Subsequence (Lcs)

The LCS (Longest Common Subsequence) method finds the longest string between two given strings that are common between the two groups and in the Here a cluster is created which mainly contain those files which are similar to the document to be compared by Longest Common Subsequence Method.Take two strings x and y and then we have to initialize a matrix of x.length *y.length. If i=0 or j=0 then the resultant value is zero. If I and j are equal then the resultant value is 1+[i-1,j-1], otherwise the resultant value is max([i-1][j],[i][j-1]). These method is used to find the longest string between the two given strings that are common between the two groups and in same order.

3. Five Modulus Method (Fmm)

The FMM (Five Modulus Method) converts a random image into 8*8 block matrix. The block matrix is divided by 5 which reduces the size of the image. Clearly, we know that each pixel is a numberbetween 0 to 255for each of the Red, Green, and Blue arrays. Therefore, if we can transform each number in that range into a number divisible by 5, then this will not affect the Human Visual System (HVS).

Mathematically speaking, any number divided by 5 will give a remainder ranges from 0-4 (e.g., 15 mod 5 is 0, 17 mod 5 is 2, 201 mod 5 is 1, 187 mod 5 is 2 and so on). Here, we have proposed a new formula to transform any number in the range 0-255 into a number that when divided by 5 the result is always lying between 0-4. Actually, any number in the range0-4 (which is the remainder of dividing 0-255 by 5) can be transformed as follows 0→(same pixel),

International Journal of Science, Engineering and Technology

An Open Access Journal

1→(-1), 2→(-2), 3→(+2), 4→(+1).We have used FMM for compression as the intended data which is required is not lost in this method. It also reduces the size of the image from higher intensity to lower intensity which helps to compare images more rapidly. FMM firstly converts the input image into gray scaled image and then compress using threshold value.

II.EXISTING SYSTEM

Plagiarism is a mistake mostly in academic field. It takes others contributions without their permission and does not give honour to the originator. Reprobates are rewarded though they are not deserved for that. We can observe plagiarism occur in various fields like literature, academic, science, music vastly.Plagiarism detection techniques are there, which are classified into character based method, structural-based method, classification or cluster based method, syntax-based methods, cross language-based methods, semantic-based methods and citation- based methods. These methods deal with plagiarism act on the text and ignore images.

A survey was carried out for plagiarism detection in 2006, which focused on text-based plagiarism detection and described some plagiarism detection tools with test cases and results. In the following year, another survey was done by Lukashenko et al., which focused on a general way of minimizes plagiarism and discussed metrics for calculating similarity scores. Although this study only showed a few attributes of seven plagiarism detection tools, this paper did not comprehensively analysis tools, detection algorithms and techniques. In 2011, Garg conducted a study that briefly discussed plagiarism and described two source code plagiarism detection tools and six natural language plagiarism detection tools.

III. DISADVANTAGES OF EXISTING SYSTEM:

- The existing system can detect the text-based plagiarism and can't be able to detect the image-based plagiarism. Images also contain a wide range of information and hence the image-based plagiarism should be detected. This is the major disadvantage of the existing system.
- As we need to compare each and every string at each instance it consumes a lot of time. Hence, it is the time consuming process.
- We can't predict the output due to the comple manual computations, these may cause errors and also affects the accuracy. The accuracy is very less due to the computational errors.

IV.PROPOSED SYSTEM

We thought about the issue involved with plagiarism detection and try to make it easy to find. We have used the LCS method to detect the plagiarism of a paper in terms of text, which not only find plagiarism but also show the percentage of plagiarism held. The images contain a very wide range of information, we have used FMM method to examine the plagiarism of a paper in terms of images.

The images contain a very wide range and especially in the computer literature to be found a lot flowchart images, the purpose of this project is to examine the plagiarism of a paper in terms of used flowchart images plagiarism using FMM. These images have been tested FMM method. The recognition accuracy average of flowchart test images that have not been tampered in terms of structure, nodes and edges in the proposed method with 81.91 percent is indicating the high success of this method and increase of recognition.

V.WORKING PRINCIPLE

In its present form, the system's primary function is to serve as a platform for training and examination. The Histogram is used during the training phase for recognition, and the network's modelling is used for testing. The analysis is based on the percentages of similarity between the photographs requested and those already present in the database.This method chooses pictures that have the highest correlations with the one you're looking for. At this stage, the expert is responsible for interpreting the findings of the correlation analysis to determine whether or not the tested images were plagiarized.

International Journal of Science, Engineering and Technology

An Open Access Journal

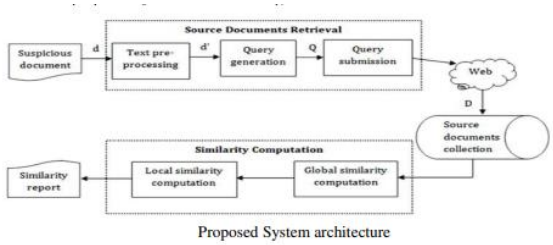

Fig.1 Proposed system Architecture.

VI.ADVANTAGES OF PROPOSED SYSTEM

- In comparison to other models that we looked at in the literature, the proposed system's accuracy is 81.91 percent, and this model ended up providing thebest accuracy.
- The output can be predicted and also consumes less time for computing, computational errors will be reduced.
- Our approach additionally includes a method which can identify the image-based plagiarism along with the text-based plagiarism.

VII.RESULTS

Our model is more accurate (82.91%) than the other models we analysed. The plagiarism is predicted more accurately in terms of both text-based and image-based.

Fig.2 Home page.

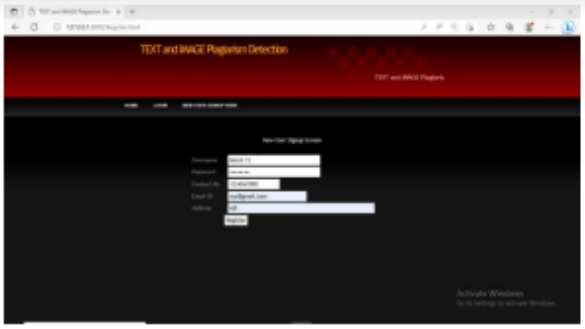

Fig.3 Signup Page.

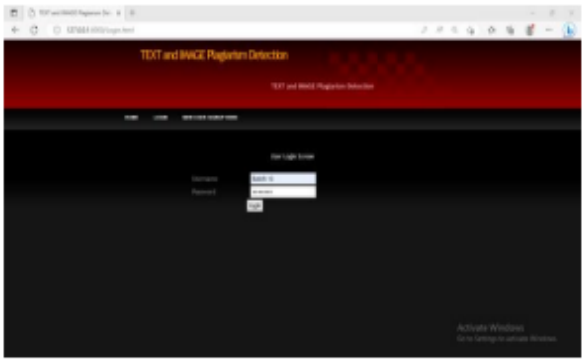

Fig.4 Login Page.

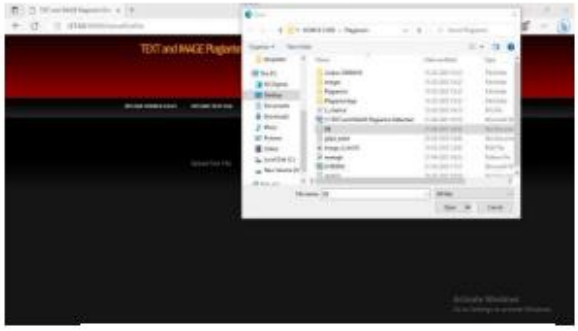

Fig.5 Upload Test File.

Fig.6 Text-based Plagiarism Detected.

Fig.7 Upload Test Image.

Fig.8 Graphs showing Image-based Plagiarism Detected.

Fig.9 Image-based Plagiarism Detected.

VIII.CONCLUSION

We introduced an image-based plagiarism detection approach that adapts itself to forms of image similarity found in academic work. The approach is achieved by including methods that analyse heterogeneous image features, selectively employing analysis methods depending on their suitability for the input image, using a flexible procedure to determine suspicious image similarities, and enabling easy inclusion of additional analysis methods in the future.

REFERENCES

1. V Lakshmi Chaitanya, "Machine Learning Based Predictive Model for Data Fusion Based Intruder Alert System", journal of algebraic statistics, Vol. 13,no. 2,pages. 2477-2483, June 2022.
2. V Lakshmi Chaitanya,'' Apriori vs Genetic algorithms for Identifying Frequent Item Sets'', International journal of Innovative Research &Development,Vol.3,no.6,pages. 249-254,June 2014.
3. Sunar mohammed Farooq,'' Static Peers for Peer-to-Peer Live Video Streaming'',Inter national journal of Scientific Engineering and Technology Research, Vol.05,No.34, Pages:7055-7064, October-2016.
4. Farooq Sunar Mahammad, P Bhaskar, A Prudvi, N Yugandhar Reddy, P Jaswanth Reddy, "Prediction Of Covid-19 Infection Based on Lifestyle Habits Employing Random Forest Algorithm", journal of algebraic statistics,Vol.13,No.3,pages.40-45,June 2022
5. Sunar Mohammed Farook, K NageswaraReddy,'' Implementation of Intrusion Detection Systems for High Performance Computing Environment Applications'', International journal of Scientific Engineering and Technology Research ,Vol.04, N0.41, Pages:8958-8963, October 2015.
6. MV Subramanyam, "Automatic feature based image registration using SIFT algorithm", conference of 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12), pages. 1-5, July 2012.
7. MV Subramanyam, Mahesh, "Feature based image registration using steerable filters and Harris algorithm",pages. 95-99, January 2012.
8. MV Subramanyam, K Satya Prasad, PV Gopi Krishna Rao,''Robust control of steam turbine system speed using improved IMC tuned PID controller'', Procedia Engineering, Vol.38,Pages. 1450- 1456,January 2012.
9. MV Subramanyam, Giri Prasad,'' A New Approach for SAR Image Denoising'', International Journal of Electrical and Computer Engineering,Vol.5,No.5, Pages. 984-991, October 2015.
10. M. Sharmila Devi, Farooq Sunar Mahammad, D. Bhavana, D. Sukanya, TV. Sai Thanusha, M. Chandrakala, P. VenkataSwathi "Machine Learning Based Classification and Clustering Analysis of Efficiency of Exercise Against Covid-19 Infection" , JOURNAL OF ALGEBRAIC STATISTICS, Vol. 13, no. 3,pages. 112-117, June 2022.
11. M. Sharmila Devi, "A comparative Study of Classification Algorithm for Printed Telugu Character Recognition", International Journal of Electronics Communication and Computer Engineering,Vol.3,no.3,pages.633-641,2012.
12. B.Swarajya Lakshmi, "Fire detection using Image processing",Asian Journal of Computer Science and Technology ISSN: 2249-0701 Vol.10 No.2, 2021, pp.14-19, 2021.
13. B.Swarajya Lakshmi, —"Identity-Based Proxy-Oriented Data Uploading and Remote Data Integrity checking in Public Cloud", International

International Journal of Science, Engineering and Technology

An Open Access Journal

Journal of Research Vol. 5, no.22,pages. 744-757, 2018.

14. SalhaAlzahrani, Vasile Palade, Naomie Salim, and Ajith Abraham. 2011. Using Structural Information and Citation Evidence to Detect Significant Plagiarism Cases in Scientific Publications. JASIST 63(2) (2011).
15. Yaniv Bernstein and Justin Zobel. 2004. A Scalable System for Identifying Coderivative Documents. In Proc. SPIRE. LNCS, Vol. 3246. Springer.
16. Teddi Fishman. 2009. "We know it when we see it"? is not good enough: toward a standard definition of plagiarism that transcends theft, fraud, and copyright. In Proc. Asia Pacific Conf. on Educational Integrity.
17. Cristian Grozea and Marius Popescu. 2011. The Encoplot Similarity Measure for Automatic Detection of Plagiarism. In Proc. PAN WS at CLEF.
18. AzharHadmi, William Puech, Brahim Ait Es Said, and Abdellah Ait Ouahman. 2012. Watermarking. Vol. 2, InTech, Chapter Perceptual Image Hashing.