ISSN 2277-2685
IJESR/Apr-Jun. 2024/ Vol-14/Issue-2/426-435
Dr.Syed Asadullah Hussaini *et. al.*, / International Journal of Engineering & Science Research

SUBJECTIVE ANSWERS EVALUATION USING MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING

Dr.Syed Asadullah Hussaini[1], Sameera Ahmedi[2], Hiba Fatima[3], Majeeda Begum[4]

[1]Associate Professor, Department of CSE, ISL College of Engineering, Hyderabad, India.

[2,3,4]B. E Student, Department of CSE, ISL College of Engineering, India.

ABSTRACT: This project proposes a novel approach that utilizes various machine learning, natural language processing techniques, to evaluate descriptive answers automatically. Solution statements and keywords are used to evaluate answers, and a machine learning model is trained to predict the grades of answers. With enough training, the machine learning model could be used as a standalone as well. Experimentation produces an accuracy of 97% with the Proposed model. Interestingly, artificial intelligence is utilized extensively as an efficient tool for predicting such a problem. The proposed work utilizes the deep learning technique along with some preprocessing steps to improve the prediction of Answer Evoluation.

INTRODUCTION

Machine Learning is a computational framework that enables algorithms to acquire knowledge and enhance their performance by analyzing examples, without requiring explicit programming by a developer. Machine learning is a component of artificial intelligence that use statistical methods to analyze data and generate predictions, which may then be utilized to derive practical insights.

The innovation lies in the concept that a computer may autonomously acquire knowledge from data, namely examples, in order to provide precise outcomes. Machine learning is intricately connected to data mining and Bayesian predictive modeling. The machine accepts data as input and employs an algorithm to generate responses.

A common objective in machine learning is to generate recommendations. Netflix suggestions are determined only by the user's history data for those who have a Netflix account. Technology businesses are using unsupervised learning techniques to enhance the user experience by tailoring recommendations.

Machine learning is used for many activities such as fraud detection, predictive maintenance, portfolio optimization, and job automation.

Machine Learning vs. Traditional Programming

Traditional programming and machine learning are distinct in their approaches and methodologies. In conventional programming, a programmer writes the code for all the regulations in collaboration with an industry specialist who specializes in the software's domain. Every rule is derived from a logical basis, and the machine will provide an output in accordance with the logical assertion. As the system becomes more intricate, further rules must be created. It may rapidly become unviable to sustain.

Traditional programming and machine learning are distinct in their approaches and methodologies. In conventional programming, a programmer writes the whole set of rules in collaboration with an industry specialist who specializes in the software's domain. Every rule is grounded in a logical basis, and the machine

will provide an output in accordance with the logical assertion. As the system becomes more intricate, further rules must be created. It may rapidly become unviable to sustain.

LITERATURE SURVEY

J. Wang and Y. Dong,, Measurement of text similarity: A survey

Data mining is a method of evaluating crime data that has been collected from several sources in order to identify patterns and trends in criminal activity. Furthermore, it may be used to enhance efficiency in expediting crime resolution and can also be employed to automatically alert authorities of criminal activity. Nevertheless, there are several data mining strategies. To enhance the effectiveness of crime detection, it is essential to carefully choose appropriate data mining approaches. This study examines the existing literature on a range of data mining applications, with a particular focus on technologies used to address criminal activities. The survey also provides insights into the research gaps and issues associated with criminal data mining. Furthermore, this study offers valuable information on the use of data mining techniques to identify patterns and trends in criminal activity. It serves as a useful resource for novice researchers interested in the field of crime data mining.

P. Patil, S. Patil, V. Miniyar, and A. Bandal ‘‘Subjective answer evaluation using machine learning

The study outlined in this document has two primary goals. One approach is to use risk terrain modeling (RTM) to predict the occurrence of shootings. The risk landscape maps generated using RTM use various contextual data pertaining to the opportunity structure of shootings to assess the likelihood of future shootings throughout a given area. The second aim was to evaluate the prediction efficacy of the risk terrain maps over two consecutive six-month intervals and to compare them with the predicting capability of retroactive hot spot maps. The findings indicate that risk terrains are a reliable predictor of future shootings, with statistical significance shown across several thresholds. Moreover, risk terrains outperform retrospective hot spot mapping in terms of accuracy. Furthermore, risk landscape maps provide simply implementable and efficient information for police administrators, enabling them to focus police patrols towards concentrated high-risk regions.

M. Faúndez-Zanuy, F. C. Morabito Italian text categorization with lemmatization and support vector machines

This study investigates a structural model of violent crime in Portland, Oregon, by analyzing the geographical distribution of both crime and its associated factors. The research utilizes conventional structural measurements derived from an opportunity framework. It presents findings using a global ordinary least squares model, which is considered to be applicable to all places within the study region. Geographically weighted regression (GWR) is subsequently shown as a viable alternative to conventional methods of crime modeling. The GWR approach employs a local model to create a collection of parameter estimates and t-values that exhibit spatial variation. Crime rates are shown to be highly influenced by various structural factors, with the strength of these associations varying depending on the area. The findings suggest that a combination model, including both spatially changing and fixed factors, may provide the most precise representation of crime. This research showcases the effectiveness of Geographically Weighted Regression (GWR) in analyzing local factors that influence crime rates and identifying errors in a global model of urban violence.

Ting, T. Washio, and G. Haffari,, ''A new simple and effective measure for bag-of-word inter-document similarity measurement

Social networks generate a vast amount of data. Twitter, a platform for microblogging, has a user base of more than 230 million active individuals who together share more than 500 million tweets on a daily basis. Our proposal involves doing an analysis of publicly available data from Twitter in order to make predictions about crime rates. The incidence of criminal activity has shown an upward trend in recent years. While crime prevention organizations use several techniques to decrease crime rates, none of the previous strategies specifically focused on analyzing the language used in Tweets to distinguish between offensive and non-offensive content as a means of predicting crime rates. This research postulates that examining the linguistic patterns in tweets might serve as a reliable indicator for forecasting crime rates in urban areas. Tweets were gathered over a 3-month period in Houston and New York City by restricting the collection to certain geographic coordinates of longitude and latitude. In addition, tweets related to criminal incidents in both locations were gathered to verify the accuracy of the prediction system. We used a Support Vector Machine (SVM) classifier to construct a predictive model for crime rates using tweets. Ultimately, we provide the verification of the prediction algorithm's accuracy in forecasting crime rates in urban areas.

M. Oghbaie and M. M. Zanjireh, 'Pairwise document similarity measure based on present term set,

This study presents a set of models that characterize the spatial and temporal patterns of criminal behavior. This argument posits that by using a simple set of processes that correlate to fundamental features in the study of crime, one may effectively monitor the emergence of hot spots. Through the examination of the most basic forms of our model, we demonstrate the emergence of a self-organized critical state of illicit behaviors, which we suggest referring to as a "warm spot" or a "tepid milieu" depending on the specific circumstances. This phenomenon is defined by a consistent presence of unlawful or uncivil behavior that remains stable without escalating, in contrast to true hotspots when localized intense activity or spikes occur. Within our framework, we conduct a more in-depth analysis of optimum policy concerns while taking into account the limitation of resources in law enforcement and deterrence. In addition, we provide enhancements to our model that include the impact of recurrent victimization, both local and long range interactions, and briefly examine the associated consequences, such as hysteresis phenomena.

PROPOSED SYSTEM:

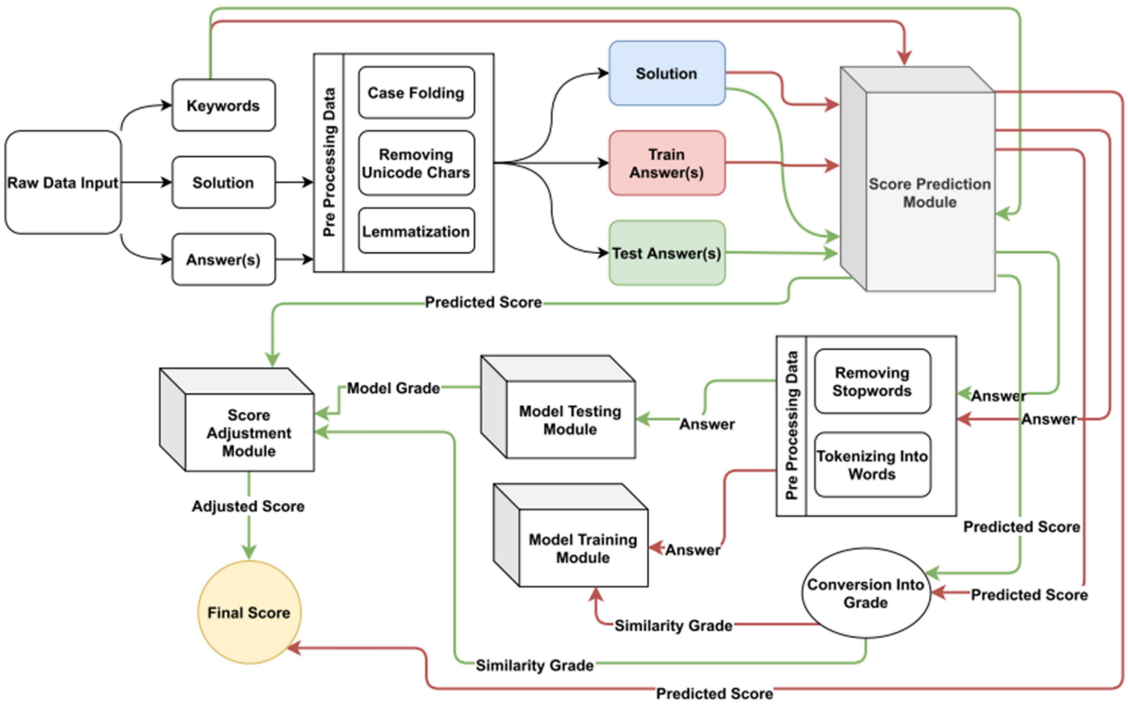
- This project proposes a new and improved way of evaluating descriptive question answers automatically using machine learning and natural language processing.
- It uses 2 step approach to solving this problem.
- First, the answers are evaluated using the solution and provided keywords using various Similarity-based techniques such as word mover's distance.
- This form of evaluation by machines is a big step forward in aiding the educational sector to perform their other duties efficiently and reduce the manual labor in trivial tasks such as comparing the answers with a correct solution.

SYSTEM ARCHITECTURE

When developing a machine learning model, it is necessary to have two datasets: one for training the model and another for assessing its performance. However, at now, we own just a solitary one. Let's divide this into two halves with a ratio of 80:20. Additionally, we will partition the dataframe into a feature column and a label column.

We have imported the train_test_split function from the sklearn library. Subsequently, use it to partition the dataset. Additionally, when test_size = 0.2, it results in a split where 80% of the data is allocated to the training dataset and 20% is allocated to the test dataset.

The random_state argument initializes the random number generator that is used to divide the dataset. The function yields four datasets. The data sets were designated as train_x, train_y, test_x, and test_y. By examining the geometry of this dataset, we can see the division or separation of the dataset. We will use a CNN Classifier, which applies several decision trees to the dataset. Ultimately, I proceed to train the model by providing the train_x and train_y data as input to the fit function.



Once the model is trained, we need to Test the model. For that we will pass *test_x* to the predict method. Random Forest is one of the most powerful methods that is used in machine learning for regression problems. The random forest comes in the category of the supervised regressor algorithm. This algorithm is carried out in two different stages the first one deals with the creation of the forest of the given dataset, and the other one deals with the prediction from the regressor.

INPUT DESIGN

The input design serves as the intermediary between the information system and the user. The process involves creating a detailed plan and set of instructions for preparing data, which is essential for making transaction data

usable for processing. This can be done by either using a computer to extract data from a written or printed document, or by manually inputting the data into the system. The input design prioritizes the management of input quantity, error control, delay avoidance, elimination of unnecessary processes, and overall simplicity of the process. The input is intentionally intended to provide both security and user-friendliness while maintaining anonymity.

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.

3. Create document, report, or other formats that contain information produced by the system.

IMPLEMENTATION

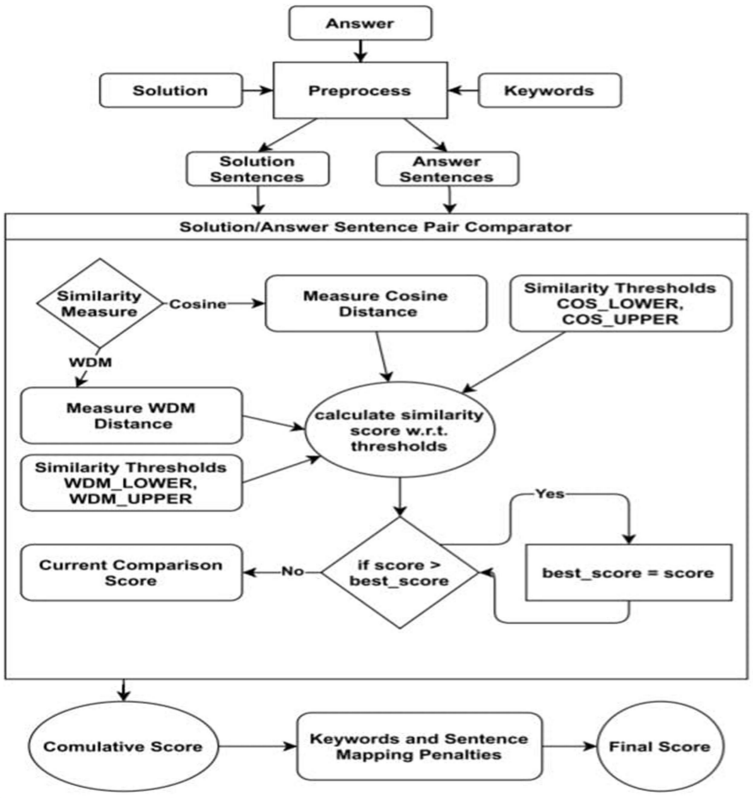
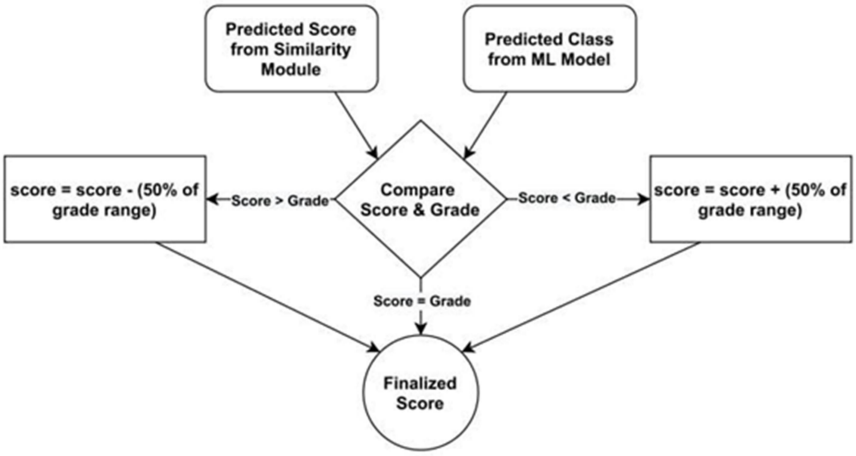


FIGURE 3. Flow chart of result prediction module.



RESULTS

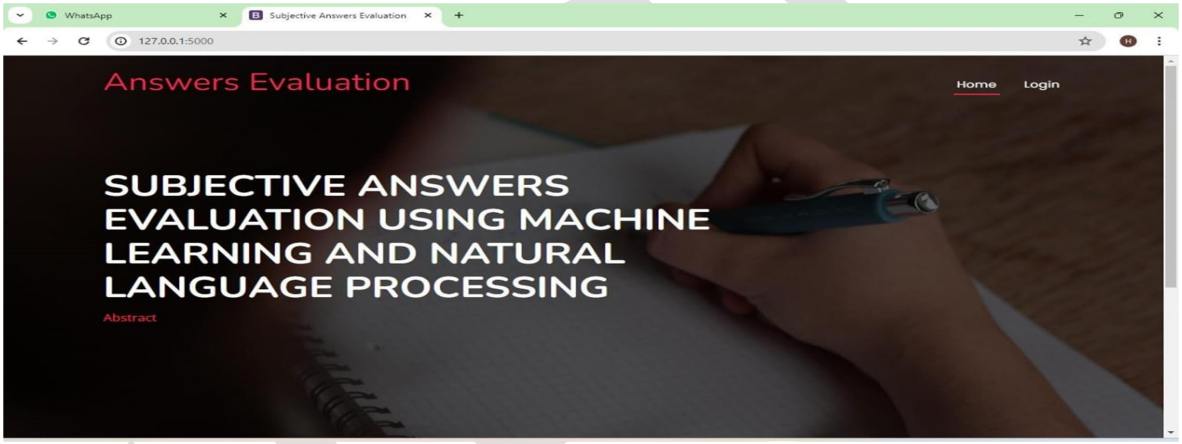


figure : Home page

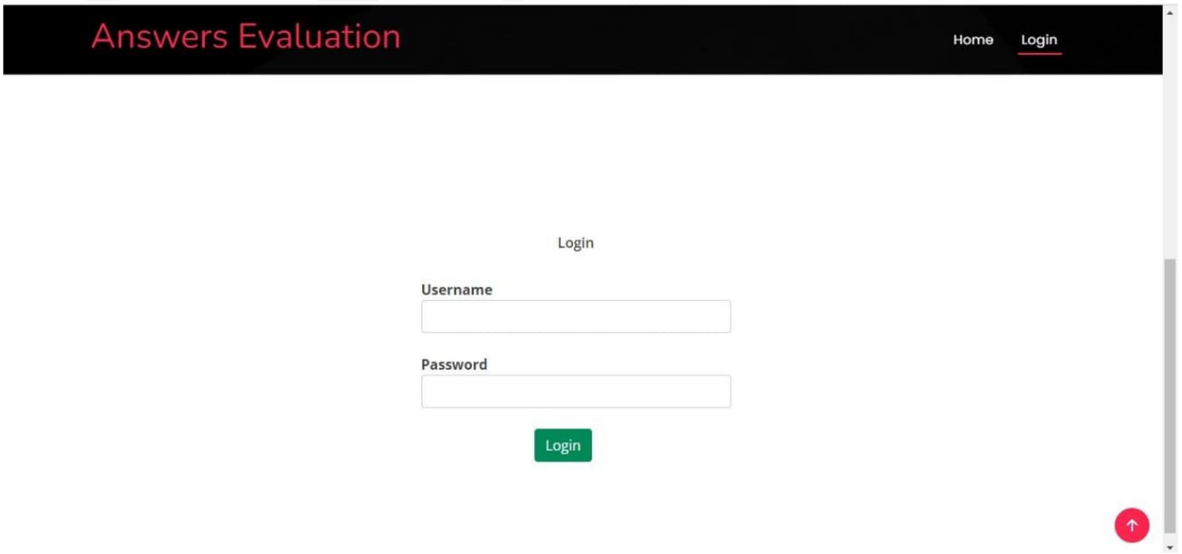


Figure : login module of admin

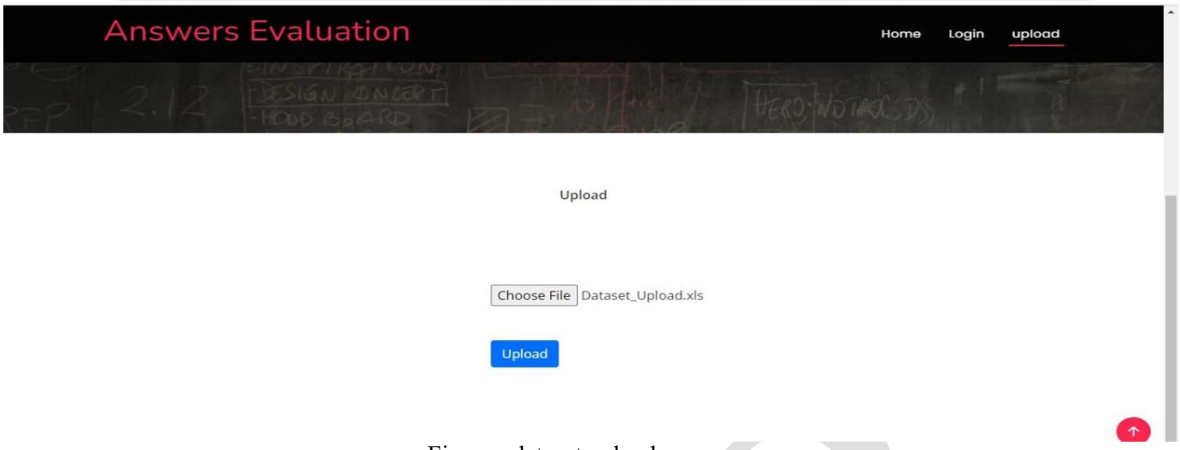


Figure : dataset upload page

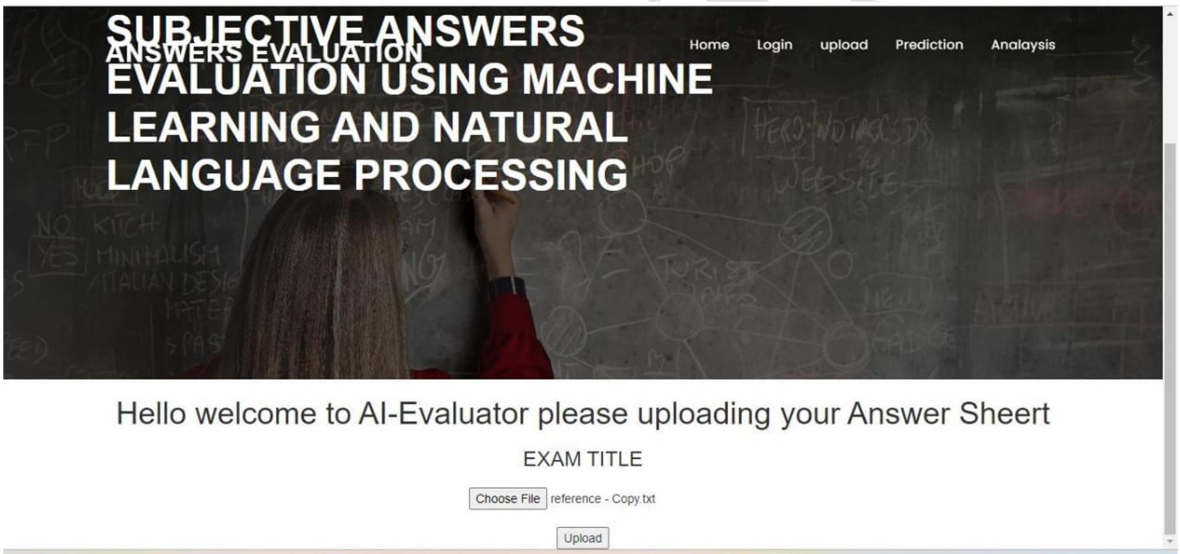


Figure : answers upload page

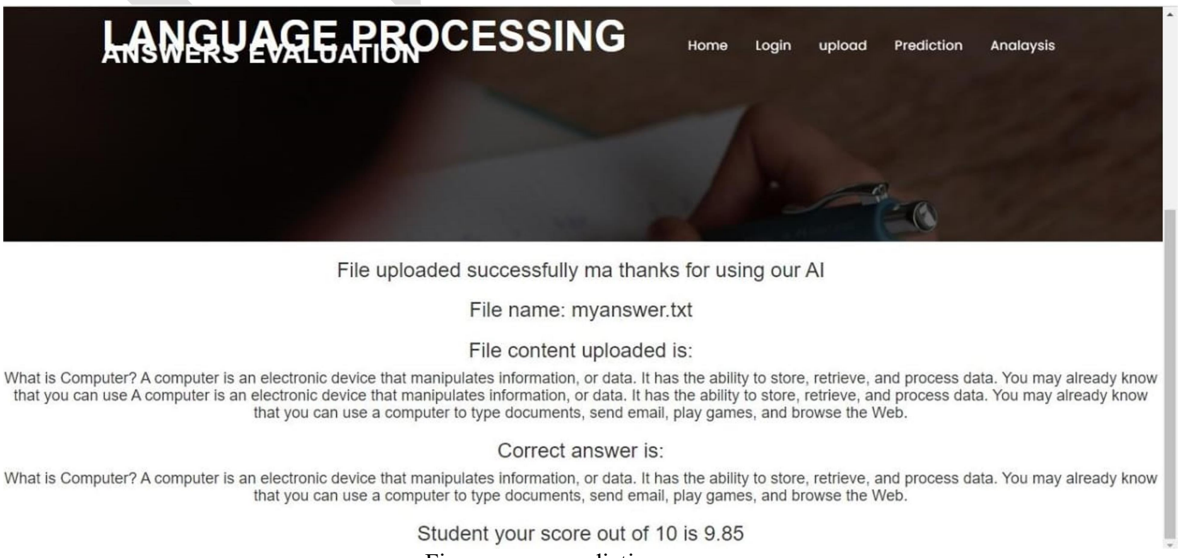


Figure : user prediction page

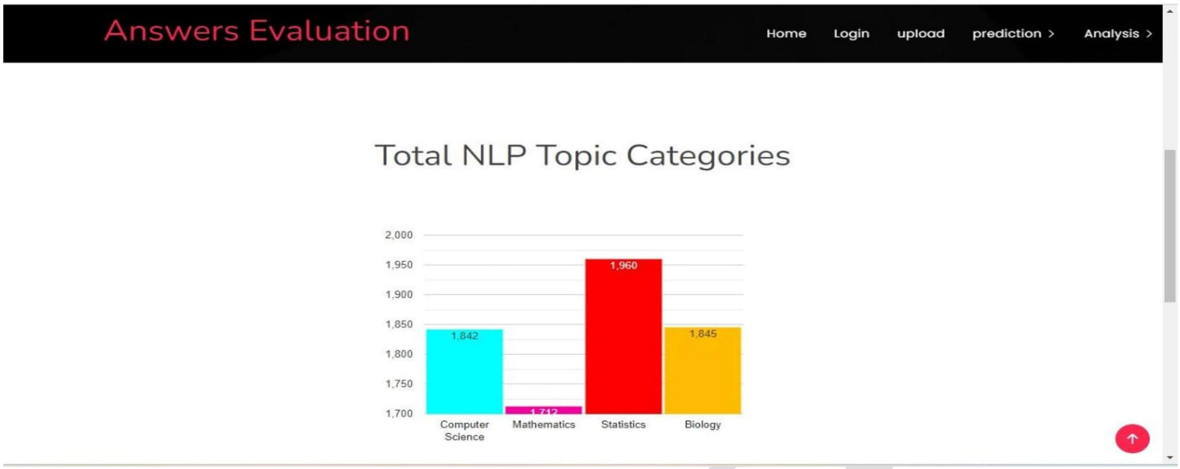
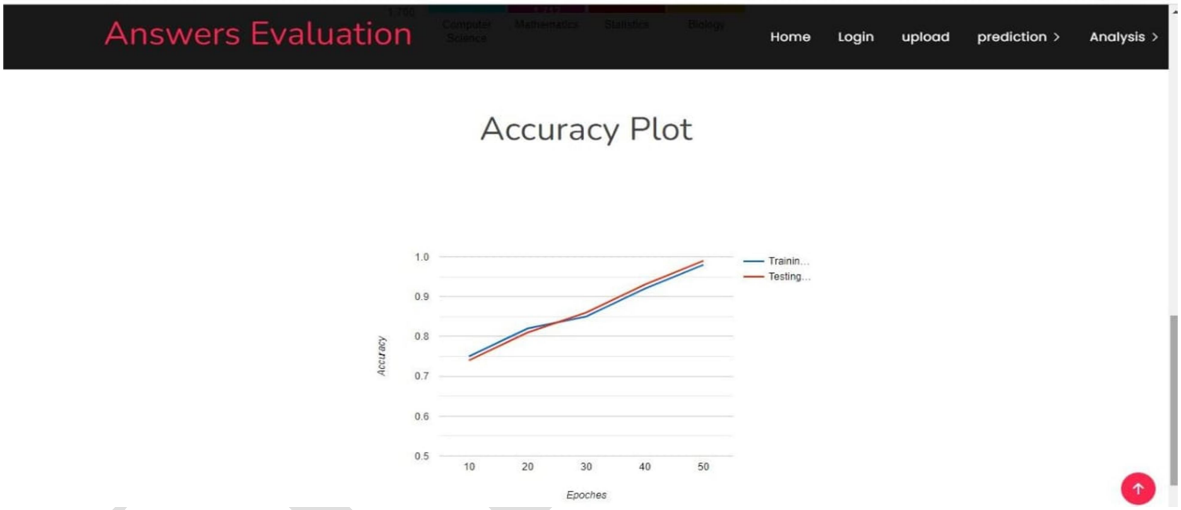


Figure : analysis page



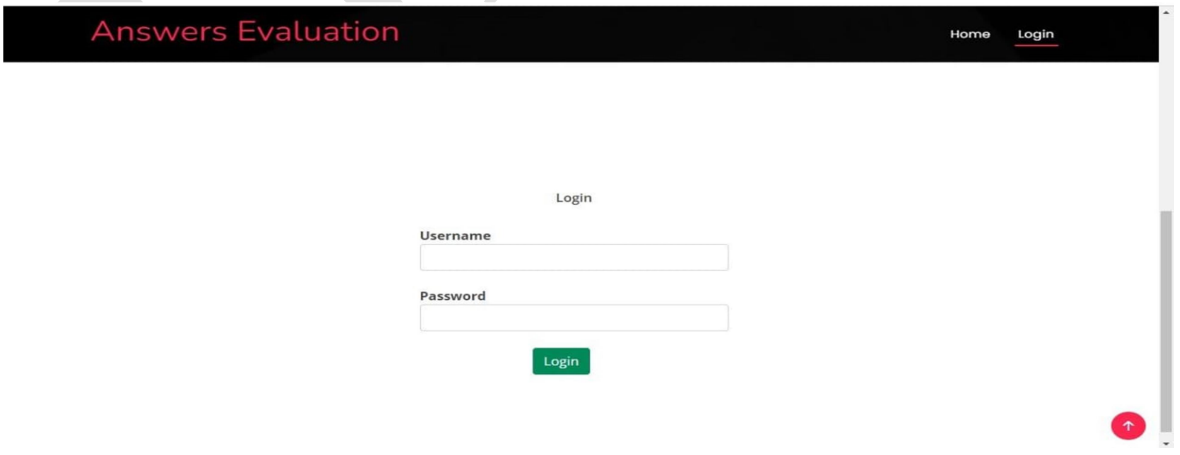


Figure : Login module of the user

This page is for verifying the user with this login option. As shown in Figure this page collects the login details of the user

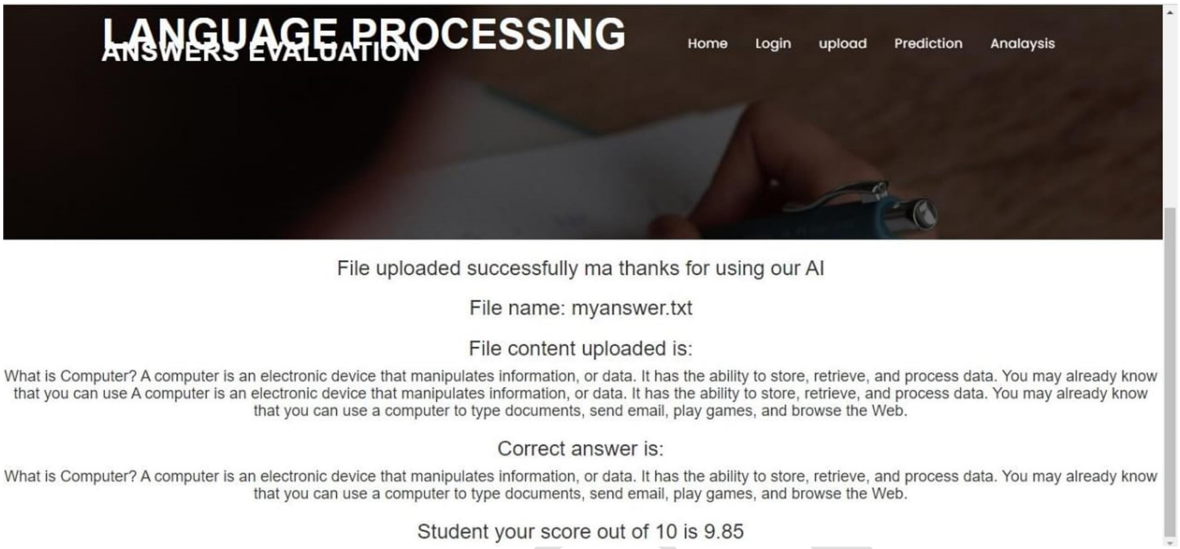


Figure : User Prediction page

As shown in Figure this page is for entering the answers of user for prediction.

The answers evaluate automatically using machine learning and natural language processing

CONCLUSION

This research presents an innovative method for evaluating subjective replies using machine learning and natural language processing methods. Two score prediction methods are suggested, yielding scores with an accuracy of up to 88%. Multiple similarity and dissimilarity criteria are examined, and many additional metrics such as the presence of keywords and the percentage mapping of sentences are used to address instances of semantically vague responses. The experimental findings demonstrate that, on average, the word2vec method outperforms existing word embedding algorithms by preserving semantic meaning. Moreover, Word Mover's Distance outperforms Cosine Similarity in the majority of circumstances and accelerates the training of the machine learning model. Through sufficient training, the model may autonomously generate predictions without relying on any semantic verification. Regarding potential enhancements, the word2vec model may be specifically trained to evaluate subjective responses in a specific area. Additionally, by using extensive data sets, the model can accommodate a much higher number of classes or grades. Assessing subjective replies is a compelling issue that we want to address. Moving forward, our goal is to discover more effective methods to resolve this challenge.

REFERENCES

[1] J. Wang and Y. Dong, ‘‘Measurement of text similarity: A survey,’’ Information, vol. 11, no. 9, p. 421, Aug. 2020.

[2] M. Han, X. Zhang, X. Yuan, J. Jiang, W. Yun, and C. Gao, ‘‘A survey on the techniques, applications, and performance of short text semantic similarity,’’ Concurrency Comput., Pract. Exper., vol. 33, no. 5, Mar. 2021.

[3] M. S. M. Patil and M. S. Patil, ‘‘Evaluating Student descriptive answers using natural language processing,’’ Int. J. Eng. Res. Technol., vol. 3, no. 3, pp. 1716–1718, 2014.

[4] P. Patil, S. Patil, V. Miniyar, and A. Bandal, ‘‘Subjective answer evaluation using machine learning,’’ Int. J. Pure Appl. Math., vol. 118, no. 24, pp. 1–13, 2018.

[5] Ijteba Sultana, Dr. Mohd Abdul Bari ,Dr. Sanjay," *Routing Performance Analysis of Infrastructure-less Wireless Networks with Intermediate Bottleneck Nodes*", International Journal of Intelligent Systems and Applications in Engineering, ISSN no: 2147-6799 IJISAE,Vol 12 issue 3, 2024, Nov 2023

[6] Md. Zainlabuddin, "*Wearable sensor-based edge computing framework for cardiac arrhythmia detection and acute stroke prediction*", Journal of Sensor, Volume2023.

[7] Md. Zainlabuddin, "*Security Enhancement in Data Propagation for Wireless Network*", Journal of Sensor, ISSN: 2237-0722 Vol. 11 No. 4 (2021).

[8] Dr MD Zainlabuddin, "*CLUSTER BASED MOBILITY MANAGEMENT ALGORITHMS FOR WIRELESS MESH NETWORKS*", Journal of Research Administration, ISSN:1539-1590 | E-ISSN:2573-7104 , Vol. 5 No. 2, (2023)

[9] Vaishnavi Lakadaram, " Content Management of Website Using Full Stack Technologies", Industrial Engineering Journal, ISSN: 0970-2555 Volume 15 Issue 11 October 2022

[10] Dr. Mohammed Abdul Bari,Arul Raj Natraj Rajgopal, Dr.P. Swetha ," *Analysing AWSDevOps CI/CD Serverless Pipeline Lambda Function's Throughput in Relation to Other Solution*", International Journal of Intelligent Systems and Applications in Engineering , JISAE, ISSN:2147-6799, Nov 2023, 12(4s), 519–526

[11] Ijteba Sultana, Mohd Abdul Bari and Sanjay," *Impact of Intermediate per Nodes on the QoS Provision in Wireless Infrastructure less Networks*", Journal of Physics: Conference Series, Conf. Ser. 1998 012029 , CONSILIO Aug 2021

[12] M.A.Bari, Sunjay Kalkal, Shahanawaj Ahamad," *A Comparative Study and Performance Analysis of Routing Algorithms*", in 3rd International Conference ICCIDM, Springer - 978-981-10-3874-7_3 Dec (2016)

[13] Mohammed Rahmat Ali,: BIOMETRIC: AN e-AUTHENTICATION SYSTEM TRENDS AND FUTURE APLLICATION", International Journal of Scientific Research in Engineering (IJSRE), Volume1, Issue 7, July 2017

[14] Mohammed Rahmat Ali,: BYOD.... A systematic approach for analyzing and visualizing the type of data and information breaches with cyber security", NEUROQUANTOLOGY, Volume20, Issue 15, November 2022

[15] Mohammed Rahmat Ali, Computer Forensics -An Introduction of New Face to the Digital World, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169-453 – 456, Volume: 5 Issue: 7

[16] Mohammed Rahmat Ali, Digital Forensics and Artificial Intelligence ...A Study, International Journal of Innovative Science and Research Technology, ISSN:2456-2165, Volume: 5 Issue:12.

[17]Mohammed Rahmat Ali, Usage of Technology in Small and Medium Scale Business, International Journal of Advanced Research in Science & Technology (IJARST), ISSN:2581-9429, Volume: 7 Issue:1, July 2020.

[18]Mohammed Rahmat Ali, Internet of Things (IOT) Basics - An Introduction to the New Digital World, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169-32-36, Volume: 5 Issue: 10

[19]Mohammed Rahmat Ali, Internet of things (IOT) and information retrieval: an introduction, International Journal of Engineering and Innovative Technology (IJEIT), ISSN: 2277-3754, Volume: 7 Issue: 4, October 2017.

[20]Mohammed Rahmat Ali, How Internet of Things (IOT) Will Affect the Future - A Study, International Journal on Future Revolution in Computer Science & Communication Engineering, ISSN: 2454-424874 – 77, Volume: 3 Issue: 10, October 2017.

[21]Mohammed Rahmat Ali, ECO Friendly Advancements in computer Science Engineering and Technology, International Journal on Scientific Research in Engineering(IJSRE), Volume: 1 Issue: 1, January 2017

[22]Ijteba Sultana, Dr. Mohd Abdul Bari ,Dr. Sanjay, "*Routing Quality of Service for Multipath Manets, International Journal of Intelligent Systems and Applications in Engineering"*, JISAE, ISSN:2147-6799, 2024, 12(5s), 08–16;

[23]Mr. Pathan Ahmed Khan, Dr. M.A Bari,: Impact Of Emergence With Robotics At Educational Institution And Emerging Challenges", International Journal of Multidisciplinary Engineering in Current Research(IJMEC), ISSN: 2456-4265, Volume 6, Issue 12, December 2021,Page 43-46

[24]Shahanawaj Ahamad, Mohammed Abdul Bari, Big Data Processing Model for Smart City Design: A Systematic Review ", VOL 2021: ISSUE 08 IS SN : 0011-9342 ;Design Engineering (Toronto) Elsevier SCI Oct : 021

[25]Syed Shehriyar Ali, Mohammed Sarfaraz Shaikh, Syed Safi Uddin, Dr. Mohammed Abdul Bari, "Saas Product Comparison and Reviews Using Nlp", Journal of Engineering Science (JES), ISSN NO:0377-9254, Vol 13, Issue 05, MAY/2022

[26] Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali," Smartphone Security and Protection Practices", International Journal of Engineering and Applied Computer Science (IJEACS) ; ISBN: 9798799755577 Volume: 03, Issue: 01, December 2021 (International Journal,U K) Pages 1-6

[27].A.Bari& Shahanawaj Ahamad, "Managing Knowledge in Development of Agile Software", in International Journal of Advanced Computer Science & Applications (IJACSA), ISSN: 2156-5570, Vol: 2, No: 4, pp: 72-76, New York, U.S.A., April 2011

[28]Imreena Ali (Ph.D), Naila Fathima, Prof. P.V.Sudha ,"Deep Learning for Large-Scale Traffic-Sign Detection and Recognition", Journal of Chemical Health Risks, ISSN:2251-6727/ JCHR (2023) 13(3), 1238-1253

[29]Imreena, Mohammed Ahmed Hussain, Mohammed Waseem Akram" An Automatic Advisor for Refactoring Software Clones Based on Machine Learning", Mathematical Statistician and Engineering ApplicationsVol. 72 No. 1 (2023)

[30] Mrs Imreena Ali Rubeena,Qudsiya Fatima Fatimunisa "Pay as You Decrypt Using FEPOD Scheme and Blockchain", Mathematical Statistician and Engineering Applications: https://doi.org/10.17762/msea.v72i1.2369 Vol. 72 No. 1 (2023)

[31] Imreena Ali , Vishnuvardhan, B.Sudhakar," Proficient Caching Intended For Virtual Machines In Cloud Computing", International Journal Of Reviews On Recent Electronics And Computer Science , ISSN 2321-5461,IJRRECS/October 2013/Volume-1/Issue-6/1481-1486

[32] Heena Yasmin, A Systematic Approach for Authentic and Integrity of Dissemination Data in Networks by Using Secure DiDrip, INTERNATIONAL JOURNAL OF PROFESSIONAL ENGINEERING STUDIES, Volume VI /Issue 5 / SEP 2016

[33] Heena Yasmin, Cyber-Attack Detection in a Network, Mathematical Statistician and Engineering Applications, ISSN:2094-0343, Vol.72 No.1(2023)

[34] Heena Yasmin, Emerging Continuous Integration Continuous Delivery (CI/CD) For Small Teams, Mathematical Statistician and Engineering Applications, ISSN:2094-0343, Vol.72 No.1(2023)