DATA MINING TECHNIQUES TO PREDICT STUDENT PERFORMANCE IN DECISION MAKING

M A Lateef Arbaz[1],.Mohammed Shahnawaz Hussain[2], Mohammed Abdul Hafeez[3], Dr. Syed Asadullah hussaini[4]

[1,2,3]B. E Student, Department of CSE, ISL College of Engineering, India.

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

ABSTRACT: An admissions system based on valid and reliable admissions criteria is very important to select candidates likely to perform well academically at institutions of higher education. This study focuses on ways to support universities in admissions decision making using data mining techniques to predict applicants' academic performance at university. A data set of 2,039 students enrolled in a Computer Science and Information College of a Saudi public university from 2016 to 2019 was used to validate the proposed methodology. The results demonstrate that applicants' early university performance can be predicted before admission based on certain pre-admission criteria (high school grade average, Scholastic Achievement Admission Test score, and General Aptitude Test score). The results also show that Scholastic Achievement Admission Test score is the pre-admission criterion that most accurately predicts future student performance. Therefore, this score should be assigned more weight in admissions systems. We also found that the Arti_cial Neural Network technique has an accuracy rate above 79%, making it superior to other classi_cation techniques considered (Decision Trees, Support Vector Machines, and Naïve Bayes).

INTRODUCTION

In the present day, all institutions of higher education, particularly those focused on computer and engineering studies, have difficulties in the admissions procedure. Every institution should aim to establish an admissions system that relies on accurate and reliable criteria to pick individuals who are likely to excel in its programs. Furthermore, it is essential for every institution to use the most effective methodologies in order to accurately forecast the prospective academic achievements of candidates prior to their admission. This would assist university decision makers in establishing efficient admissions standards. Nevertheless, the majority of higher education institutions have difficulties when they evaluate their extensive educational datasets in order to forecast students' academic success [1]. The reason for this is because they rely only on traditional statistical approaches instead of using more advanced and effective prediction techniques like Educational Data Mining (EDM), which is now the most widely used method for assessing and forecasting students' academic performance [2]_[6]. Educational Data Mining (EDM) involves collecting valuable information and trends from a vast educational database [2], which may then be used to forecast students' academic success. Due to enhanced knowledge, student performance may be significantly enhanced via more efficient strategic plans. This research aims to assist colleges in their admissions process by using data mining methods to accurately forecast the academic performance of candidates prior to admission.

LITERATURE SURVEY

[1] The paper titled "Data Mining Approach for Predicting Student and Institution's Placement Percentage" was presented by Professor Ashok M and Assistant Professor Apoorva A at the 2016 International Conference on Computational Systems and Information Systems for Sustainable Solutions. The author of this study has used the data mining approach to forecast the placement of students. To forecast student placement, the author has separated the data into two parts. The first section is the training segment, which consists of historical data of students who have already graduated. Another section has the most recent data of students. Using the historical data, the author has developed an algorithm to calculate the likelihood of placement. The author used a range of data mining methods, including decision tree, Naive Bayes, and neural network. Additionally, a proposed algorithm was utilized. The decision-making process was facilitated by the usage of a confusion matrix.
[2] The paper titled "Student Placement Analyzer: A Recommendation System Using Machine Learning" was authored by Senthil Kumar Thangavel, Divya Bharathi P, and Abijith Sankar. It was presented at the International Conference on Advanced Computing and Communication Systems (ICACCS -2017) held on January 6-7, 2017, in Coimbatore, India. This study focuses on the issues that institutes encounter in relation to placement. Forecasting the placement outcomes becomes more complex as the number of organizations inside an institution grows. Machine learning can effectively address this intricate prediction issue. This paper considers the whole academic record of the student. Multiple classification and data processing methods, such as Naïve Bayes, Decision Tree, Support Vector Machine (SVM), and Regression, are used. Once the students' predictions are made, they might be categorized as core companies, dream companies, or support services.
[3] Giri, A., Bhagavath, M. V. V., Pruthvi, B., Dubey, N. (2016). "A Placement Prediction System Using K-Nearest Neighbors Classifier". In Second International Conference on Cognitive Computing and Information Processing (CCIP). The placement prediction algorithm utilizes K-Nearest Neighbors categorization to estimate the likelihood of students being hired by different firms. The resulting result is also contrasted with the results produced from other machine learning models such as Logistic Regression and Support Vector Machines (SVM). Companies evaluate students' academic histories and skill sets, including programming abilities, communication skills, analytical skills, and teamwork, throughout the recruiting process. This system collects data from the two most recent batches.
[4] In the paper titled "Class Result Prediction using Machine Learning", the authors Pushpa S K, Associate Professor, Manjunath T N, Professor and Head, Mrunal T V, Amartya Singh, and C Suhas presented their research at the International Conference On Smart Technology for Smart Nation in 2017. This research use machine learning to predict the outcome of a class. The performance of students in the previous semester, as well as their results in the internal exams of the current semester, are taken into account to determine whether they will pass or fail in the current semester before taking the final test. The author used Support Vector Machines (SVM), Naive Bayes, Random Forest Classifier, and Gradient Boosting algorithms to calculate the outcome. Boosting is a technique in ensemble learning that combines many learning algorithms to achieve improved predicting performance.
[5] "Student Placement Analyzer: A Recommendation System Utilizing Machine Learning" by Apoorva Rao R, Deeksha K C, Vishal Prajwal R, Vrushak K, Nandini, JARIIE-ISSN(O)-2395-4396 Institutions now have several issues in relation to student placements. Managing and forecasting student placement manually is a challenging endeavor for educational organizations. In order to address these issues, researchers have delved

into the notion of machine learning and numerous algorithms to forecast the outcomes of students in a class. The training data set consists of historical data from previous students, which is used to train the model. This software method accurately forecasts the placement status in five distinct categories, namely dream company, core company, mass recruiter, not eligible, and not interested in placements. This technique is beneficial for pupils who are academically inferior. Institutions might provide additional support to students who are academically inferior in order to enhance their performance. The Naïve Bayes algorithm will monitor all the data and give suitable decisions.

IMPLEMENTATION

Modules:

Dataset Upload:

In this module, administrator can upload the student training dataset. For this, admin should have to select only .xlsx files only. But if any other file format uploaded by administrator then this system can return error message.

View Dataset:

This module can display the student dataset in table format. Here the administrator can monitor the training dataset which is uploaded properly or not. This result tables can be display with column as attribute names and rows as attribute values.

Performance Evaluations:

This module can shows the experimental results of four machine learning algorithms. It can display the bar chart graph with comparisons of accuracy of four machine learning techniques. This module can be decided which classifier is providing good accuracy compare to remaining algorithms. In our system, neural network classifier is giving best accuracy.

Prediction:

This module will provide the input form to predict the student performance for admission of university. This module can be operating by user. Here user will get input form which is have subject marks then the user needs to fill the all fields then this system can return the student performance results as university name by performing the prediction with neural network classifier.

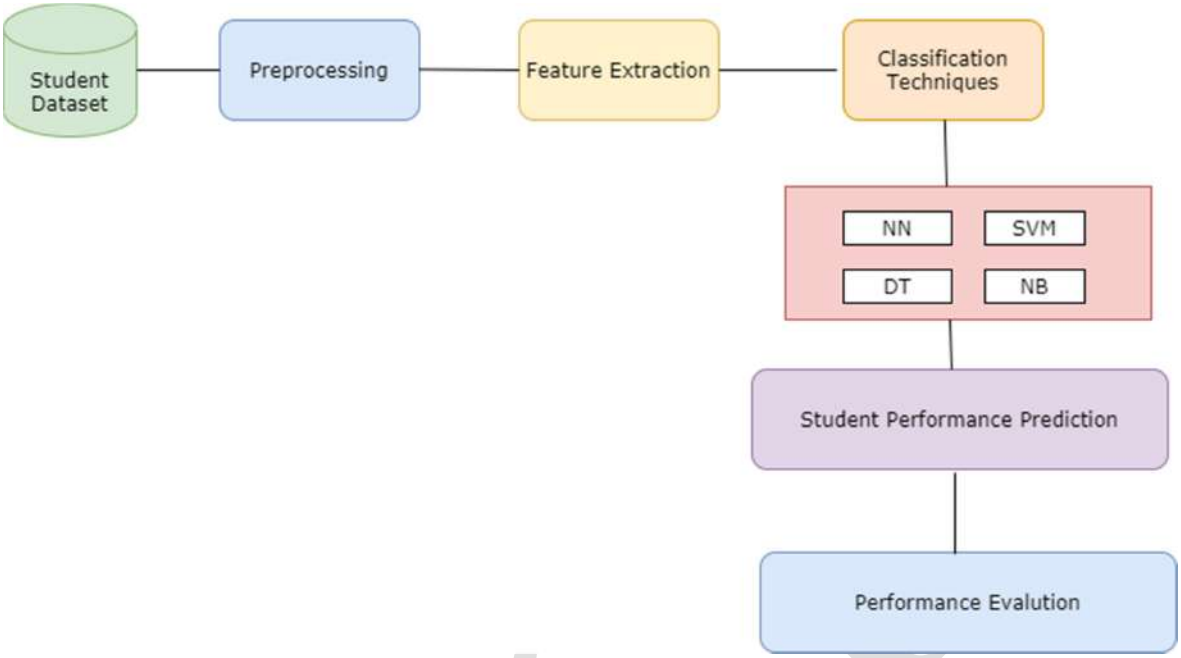
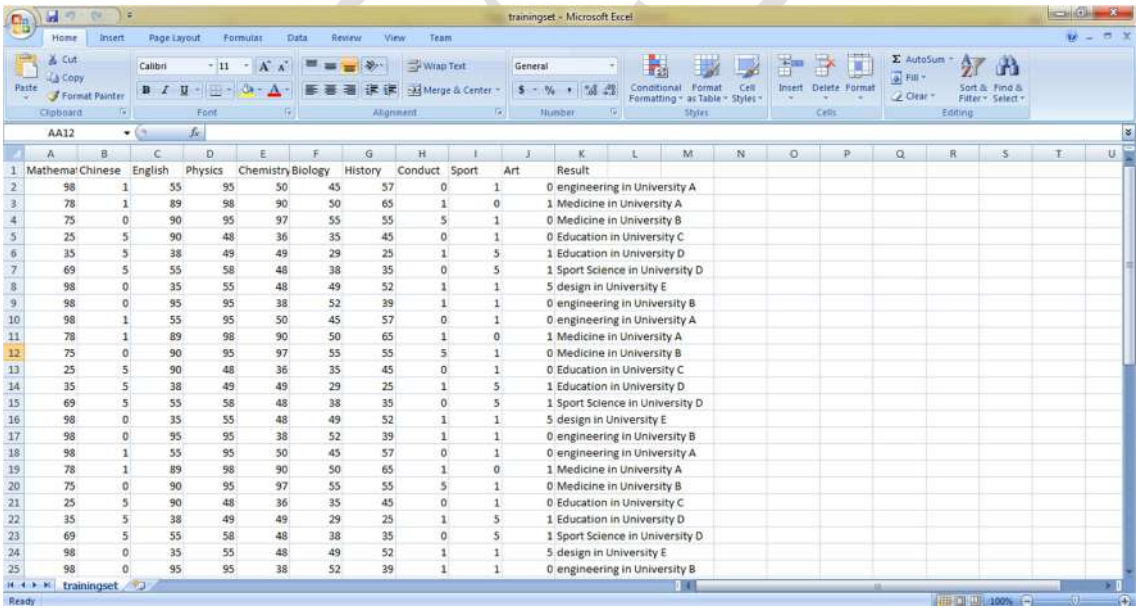


Figure.1 Student dataset

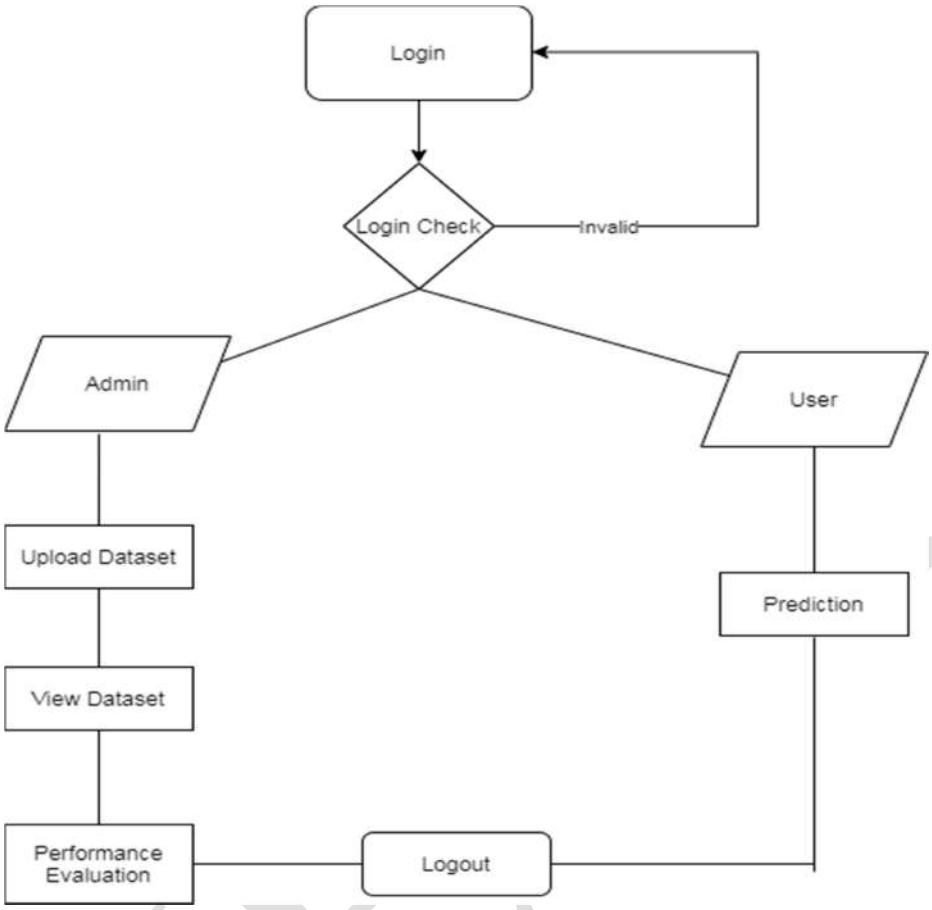


trainingset - Microsoft Excel

	Mathema	Chinese	English	Physics	Chemistry	Biology	History	Conduct	Sport	Art	Result
2	98	1	55	95	50	45	57	0	1	0	engineering in University A
3	78	1	89	98	90	50	65	1	0	1	Medicine in University A
4	75	0	90	95	97	55	55	5	1	0	Medicine in University B
5	25	5	90	48	36	35	45	0	1	0	Education in University C
6	35	5	38	49	49	29	25	1	5	1	Education in University D
7	69	5	55	58	48	38	35	0	5	1	Sport Science in University D
8	98	0	35	55	48	49	52	1	1	5	design in University E
9	98	0	95	95	38	52	39	1	1	0	engineering in University B
10	98	1	55	95	50	45	57	0	1	0	engineering in University A
11	78	1	89	98	90	50	65	1	0	1	Medicine in University A
12	75	0	90	95	97	55	55	5	1	0	Medicine in University B
13	25	5	90	48	36	35	45	0	1	0	Education in University C
14	35	5	38	49	49	29	25	1	5	1	Education in University D
15	69	5	55	58	48	38	35	0	5	1	Sport Science in University D
16	98	0	35	55	48	49	52	1	1	5	design in University E
17	98	0	95	95	38	52	39	1	1	0	engineering in University B
18	98	1	55	95	50	45	57	0	1	0	engineering in University A
19	78	1	89	98	90	50	65	1	0	1	Medicine in University A
20	75	0	90	95	97	55	55	5	1	0	Medicine in University B
21	25	5	90	48	36	35	45	0	1	0	Education in University C
22	35	5	38	49	49	29	25	1	5	1	Education in University D
23	69	5	55	58	48	38	35	0	5	1	Sport Science in University D
24	98	0	35	55	48	49	52	1	1	5	design in University E
25	98	0	95	95	38	52	39	1	1	0	engineering in University B

Figure.1 Student dataset

DFD:



MANUAL TESTING ON PROJECT APPLICATION

TEST CASES

Test Case ID #1		**Test Case Description** - Validations in Registration Form		
S#	**Prerequisites**	**S#**	**Test Data Requirement**	
1	User should be Registered	1	Data should be valid	
Test Condition				
Entering data in registration form				
Step #	**Step Details**	**Expected Results**	**Actual Results**	**Pass/Fail/Not Executed/Suspended**
1	User gives First and Last Name	Pop showing email verification message	Enter valid email/password	Fail
2	Submitting the form without entering any details	Pop showing email verification message	Enter email /password	Fail
3	User enters invalid	Pop showing	Enter valid email id	Fail

	format of email id	email verification message		
4	User enters a phone number with < 10 digits	Pop showing email verification message	Enter valid phone number	Fail
5	Entering valid username and password	Pop showing email verification message	Pop showing email verification message	Pass

Table 1 Registration test case

Test Case ID #2		**Test Case Description** - Validations in Login Form		
S#	**Prerequisites**	**S#**	**Test Data Requirement**	
1	User should have an email id	1	Data should be valid	
Test Condition				
Entering data in login form				
Step #	**Step Details**	**Expected Results**	**Actual Results**	**Pass/Fail/Not Executed/Suspended**
1	User gives aemail or password of <6 characters	User logged in	Enter valid email/password	Fail
2	Submitting the form without entering any details	User logged in	Enter email /password	Fail
3	User enters wrong Email and (or) password	User logged in	Enter correct email /password	Fail

Table 2 Login test case

Results

Result of My Application on UC browser

Result of my Project in chrome

Result of my Project in Opera

Checking on Different Browsers out all the major compatibility issues, you need to perform a round of cross-browser testing in general on minor but important factors like

Browser	Compatibility issues	Result
Chrome	Alignment of elements	YES
	Pop-Ups	YES
	Alignment of checkboxes	YES
	Alignment and functioning of buttons	YES
	URL redirection from buttons	YES
	Drop-down Menus	YES
	Forms and Form APIs	YES
	Grids/Tables	YES
	Sessions and cookies	YES
	Dates	YES
	Zoom in and Zoom out functionality	YES
	Appearance of scroll	YES
	Flash	YES
	HTML animations	YES
	Mouse hovering	YES
	Image alignment	YES
	Alt tags	YES

Test results on chrome

Browser	Compatibility issues	Result
Opera	Alignment of elements	YES
	Pop-Ups	YES
	Alignment of checkboxes	YES
	Alignment and functioning of buttons	YES
	URL redirection from buttons	YES
	Drop-down Menus	YES
	Forms and Form APIs	YES
	Grids/Tables	YES
	Sessions and cookies	YES
	Dates	YES
	Zoom in and Zoom out functionality	YES
	Appearance of scroll	YES
	Flash	YES

	HTML animations	YES
	Mouse hovering	YES
	Image alignment	YES
	Alt tags	YES

Test Results on opera

Browser	Compatibility issues	Result
Internet exp	Alignment of elements	YES
	Pop-Ups	YES
	Alignment of checkboxes	YES
	Alignment and functioning of buttons	YES
	URL redirection from buttons	YES
	Drop-down Menus	YES
	Forms and Form APIs	YES
	Grids/Tables	YES
	Sessions and cookies	YES
	Dates	YES
	Zoom in and Zoom out functionality	YES
	Appearance of scroll	YES
	Flash	YES
	HTML animations	YES
	Mouse hovering	YES
	Image alignment	YES
	Alt tags	YES

Test results on internet exp

Browser	Compatibility issues	Result
UC Browser	Alignment of elements	YES
	Pop-Ups	YES
	Alignment of checkboxes	YES
	Alignment and functioning of buttons	YES
	URL redirection from buttons	YES
	Drop-down Menus	YES
	Forms and Form APIs	YES
	Grids/Tables	YES
	Sessions and cookies	YES
	Dates	YES

	Zoom in and Zoom out functionality	YES
	Appearance of scroll	YES
	Flash	YES
	HTML animations	YES
	Mouse hovering	YES
	Image alignment	YES
	Alt tags	YES

Test results on UC Browser

Browser	Compatibility issues	Result
Edge Browser	Alignment of elements	YES
	Pop-Ups	YES
	Alignment of checkboxes	YES
	Alignment and functioning of buttons	YES
	URL redirection from buttons	YES
	Drop-down Menus	YES
	Forms and Form APIs	YES
	Grids/Tables	YES
	Sessions and cookies	YES
	Dates	YES
	Zoom in and Zoom out functionality	YES
	Appearance of scroll	YES
	Flash	YES
	HTML animations	YES
	Mouse hovering	YES
	Image alignment	YES
	Alt tags	YES

Test results on Edge Browser

CONCLUSION

The aim of this study is to support higher education institutions in making good decisions in its admissions process by predicting applicants' academic performance before admitting them. Four prediction models were proposed and developed using four well-known data mining techniques, namely: Articial Neural Network (ANN), Decision Tree, Support Vector Machine (SVM), and Naive Bayes. The study was conducted with a dataset of 2,039 records of students enrolled in PNU, one of the largest universities in KSA. The methods used, however, are general and can be used in any higher education institution. The study conrms the effectiveness of

prediction modeling in higher education institutions where decision makers can use these models in planning and optimizing institutions' limited resource allocations.

REFERENCES

[1] [1] H. Guruler, A. Istanbullu, and M. Karahasan, ``A new student performance analysing system using knowledge discovery in higher educational databases," Comput. Edu., vol. 55, no. 1, pp. 247254, Aug. 2010.

[2] [2] S. K. Mohamad and Z. Tasir, ``Educational data mining: A review," Procedia Social Behav. Sci., vol. 97, pp. 320324, Nov. 2013.

[3] [3] A. Peña-Ayala, ``Educational data mining: A survey and a data miningbased analysis of recent works," Expert Syst. Appl., vol. 41, no. 4, pp. 14321462, Mar. 2014.

[4] [4] C. Romero and S. Ventura, ``Educational data mining: A review of the state of the art," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 6, pp. 601618, Nov. 2010. \

[5] [5] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, ``Educational data mining and learning analytics for 21st century higher education: A review and synthesis," Telematics Informat., vol. 37, pp. 1349, Apr. 2019.

[6] [6] C. Anuradha and T. Velmurugan, ``A comparative analysis on the evaluation of classication algorithms in the prediction of students performance," Indian J. Sci. Technol., vol. 8, no. 15, pp. 9746846, Jan. 2015.

[7] 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

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

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

[10] 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)

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

[12] 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

[13] 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

[14] 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)

[15] 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

[16] 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

[17] 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

[18] 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.

[19] 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.

[20] 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

[21] 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.

[22] 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.

[23] 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

[24] 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;

[25] 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

[26] 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

[27] 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

[28] 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

[29] .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

[30] 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

[31] 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)

[32] 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)

[33] 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

[34] 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

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

[36] 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)