Research Inventy: International Journal of Engineering And Science
Vol.13, Issue 11 (November 2023), PP 25-30
Issn (e): 2278-4721, Issn (p):2319-6483, www.researchinventy.com

Data Representation, Features Extraction, Assessment and Prediction of ECE Admission in ACET

Dr Dola Sanjay S[1]
[1] *Professor, Aditya College of Engineering and Technology, Kakinada, Andhra Pradesh, India.*
[1]*Post Doctorial Fellowship, Central Christian University, East Africa, Malawi*

Abstract:
Information about admission data is pre processed and analyzed. The ECE branch intake sanctioned as per AICTE and number of students admitted year wise are discussed for two decades. Various data tools are used for data processing, methods and representations of data sets are assessed and validated. Our ability to detect, to understand, and to address student difficulties is highly dependent on the capabilities of the tool. Feedback from numerous sources has considerably improved the educational materials, which is a continuing task.
Keywords: *Python, MATLAB, ECE, prediction, branch, data*

Date of Submission: 06-11-2023 Date of acceptance: 20-11-2023

I. INTRODUCTION

1.1 Data Acquisition and Extracting the Features

1.1.1 Preprocessing student database

Preprocessing and creating raw data sets of student data and segmentation process is performed. As mentioned earlier, there are two types of data sets: 1) Educational resources such as web pages, demonstrations, simulations, and individualized problems designed for use on homework assignments, quizzes, and examinations; 2) Information about users, who create, modify, assess, or use these resources.
The original data are stored with escape sequence codes as shown in Figure 2.1.1:

From the Fig 2.1.1(a) and Fig 2.1.1(b) it shows the trends in intake and admitted strength of ECE branch for two decades, this branch was started in the year 2004 with an intake of 60, as there was a good demand for ECE branch, the intake has been increased to 120, followed
by 180 and upto 240, due to high demand in CSE and CSE related allied branches, the intake has been reduced to 180 from the academic year 2022.

Table 3.1.2 is providing the trends in two decades of admitted students both gender wise and social status wise, if we could observe, the social status ST takers are very low; it can be treated as no takers, when compared to other social status.

II. EXPERIMENTAL PROCEDURE

The student data restored from .db files from a student directory and fetched into a hash table. The special hash keys "*keys*", "*version*" and "*timestamp*" were obtained fromthe hash. The *version* will be equal to the total number of versions of the data that have been stored. The *timestamp* attribute is the UNIX time the data was stored. *keys* is available in every historical section to list which keys were added or changed at a specific historical revision of a hash. We extract some of the features from a structured homeworkdata, which is stored as particular URL's.

2.1.1 Preprocessing Activity Log

ECE-CAPA records and dynamically organizes a vast amount of information on students' interaction with and understanding of these materials. Since ECE-CAPA logs every activity of every student who has used online

Table 2.1.1: Total Admission gender wise for two decades

	INTAKE	ADMITTED	PW	SC		ST		BC-A		BC-B		BC-C		BC-D		BC-E		OC		TOTAL	
Year				Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
2004-2005	60	60			1					2	2	1		3	5			26	20	32	28
2005-2006	90	90		2	1			4	4	5	3	1	1	5	2	1		38	23	56	34
2006-2007	90	90		3	1			3	5	3	5	1	1	7	4	2		32	23	51	39
2007-2008	90	90		1				6	4	12	5			3	3			35	21	57	33
2008-2009	120	130		14	6			3	4	9	4		1	7	4	2	1	53	22	88	42
2009-2010	120	120		9	7	1		6	3	9	3			6	7	3		44	22	78	42
2010-2011	120	120		9	10	1		4	5	7	6		1	5	5	1		38	28	65	55
2011-2012	180	180		6	1			7	9	7	17	2	1	11	9	4		59	47	96	84
2012-2013	240	240		10	9			9	4	15	22		2	11	8	2	3	73	72	120	120
2013-2014	240	191		3	2				2	20	13	1		11	9			68	62	103	88
2014-2015	240	201		2	1			2	1	18	10	1		6	8	1		80	70	110	90
2015-2016	240	240		4	2			10	5	26	21		3	11	13	1	1	66	77	118	122
2016-2017	240	239		2	1			16	2	21	19	2		15	13			67	81	123	116
2017-2018	240	237			1		1	8	3	27	22	2		24	8	1	2	68	70	130	107
2018-2019	240	234		3	2			2	1	28	15		1	20	4		1	86	69	139	93
2019-2020	257	247	257	3				12	3	27	19			20	13			73	77	135	112
2020-2021	257	241	257	9	3	1		19	6	22	17		2	18	17	2		56	69	127	114
2021-2022	240	262		34	8			13	9	26	20	2		24	13	4		53	56	156	106
2022-2023	180	191		9	4		1	13	6	19	13			11	7	2	1	53	52	107	84

educational resources and their recorded paths, the activity.log usually grows faster when students have more access to the educational resources. A sample of different types of data, which are logged in activity.log after a preprocessing phase, is shown in figure 2.1.1.

2.1.2 Extractable Features

An essential step to perform classification is selecting the features used for classification. The following features are examples of those stored by the ECE-CAPA system:

- Total number of correct answers.
- Getting the problem right on the first try.
- Number of attempts before correct answer is derived.
- Total time that passed from the first attempt, until the correct solution was demonstrated, regardless of the time spent logged in to the system. Also, the time at which the student got the problem correct relative to the due date.
- Total time spent on the problem regardless of whether they got the correct answer or not. Total time that passed from the first attempt through subsequent attempts until the last submission was demonstrated.
- Participating in the communication mechanisms, versus those working alone.
  ECE-CAPA provides online interaction both with other students and with the instructor.
- Reading the supporting material before attempting homework vs. attempting the homework first and then reading up on it.
- Submitting a lot of attempts in a short amount of time without looking up material in between, versus those giving it one try, reading explanatory/supportive material, submitting another one, and so forth.
- Giving up on a problem versus students who continued trying up to the deadline.
- Time of the first log on (beginning of assignment, middle of the week, lastminute) correlated with the number of submissions or number of solved problems.

These features enable ECE-CAPA to provide many assessments tools for instructors as itwill be explained in the next section, by 180 and upto 240, due to high demand in CSE and CSE related allied branches, the intake has been reduced to 180 from the academic year 2022.

From the fig 2.1.1(b), the number of admitted students between the years 2012 and 2015 has a slight drop, during these 4 years, before the improvement of admitted students in 2016

all students, sorted according to the problem order. In this step, ECE-CAPA has provided in the following statistical information:

1. #Stdnts: Total number of students who take a look at the problem.(Let #Stdnts is equal to n)
2. Tries: Total number of submissions to solve the problem of a student.
3. Mod: Mode, maximum number of submissions for solving the problem.
4. Mean: Average number of the submissions.
5. #YES: Number of students solved the problem correctly.
6. #yes: Number of students solved the problem by override.

   Sometimes, a student gets a correct answer after talking with the instructor. This type of correct answer is called “corrected by override".
7. %Wrng: Percentage of students tried to solve the problem but still incorrect.

8. S.D.: *Standard Deviation* of the students' submissions.

$$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

9. Skew.: *Skewness* of the students' submissions.
10. DoDiff: *Degree of Difficulty* of the problem.

Clearly, the Degree of Difficulty is always between 0 and 1. This is a useful factor for an instructor to determine whether a problem is difficult, and the degree of this difficulty. Thus, DoDiff of each problem is saved in its meta data.

III. RESULTS AND DISCUSSIONS

3.1 Feedback to the instructor from online homework

ECE-CAPA has enabled instructors to efficiently create and distribute a wide variety of educational materials, assignments, assessments, etc. These include numerous types of formative conceptual and algorithmic exercises for which prompt feedback and assistance can be provided to students as they work on assigned tasks. This section presents recent developments that allow rapid interpretation of such data in identifying students' misconceptions and other areas of difficulty, so that concurrent or timely corrective action can be taken. This information also facilitates detailed studies of the educational resources used and can lead to redesign of both the materials and the course.

3.1.1 Student Evaluation

An important task of the feedback tools for the instructor is to help identify the source of difficulties and the misconceptions students have about a topic. There are basically three ways to look at such homework data: by student, by problem, or cross- cutting (per student, per problem).

The amount of data gathered from large enrollment courses (200-400 students) with over 200 randomizing homework problems, each of them allowing multiple attempts, can be overwhelming. Figure 3.1.1.1 shows just a small excerpt of the homework performance in an introductory physics course, students in the rows, problems in the columns, each character representing one online homework problem for one student. A number shown is the number of attempts it took that particular student to get that particular problem correct ("*" means more than nine attempts).

1..9: correct by student in 1..9 submissions	*: correct by student in more than 9 submissions
+: correct by override	-: incorrect by override
.: incorrect attempted	#: ungraded attempted
' ': not attempted	x: excused

Figure 3.1.1.1 A small excerpt of the performance overview for a small introductory physics class

We extract from student data some reports of the current educational situation of every student as shown in table 3.1.1.1. A 'Y' shows that the student has solved the problem and an 'N' shows a failure. A '-' denotes an un-attempted problem. The numbers in the right column show the total number of submissions of the student in solving the corresponding problems.

For a per-student view, each of the items in the table in 3.2.1.1 is clickable and shows both the students' version of the problem (since each is different), and their previous attempts. Figure is an example of this view, and indicates that in the presence of a medium between the charges, the student was convinced that the force would increase, but also that this statement was the one he was most unsure about: His first answer was that the force would double; no additional feedback except "incorrect" was provided by the system. In his next attempt, he would change his answer on only this one statement (indicating that he was convinced of his other answers) to "four times the force" – however, only ten seconds passed between the attempts, showing that he was merely guessing by which factor the force increased.

3.1.2 Conceptual Problems

An important task of the feedback tools for the instructor is to help identify the source of difficulty in numerical algorithmic questions, but it also allows for the identification of misconceptions students may have on qualitative questions. Student responses to two qualitative exercises, one from physics and the second from vector math, illustrate the way that the analysis tool detects difficulties and their source, specific misconceptions. The physics question is Problem 14 from assignment 8, which as indicated above, had five days before it was due. As shown in Figure 3.9 that problem averaged at that time slightly more than 4 submissions per successful solution. There were 50 correct solutions as a result of 208 submissions by 74 students. The order in which the six statements are presented varies among students. Each statement is selected randomly from one of the six concept groups. Each concept group focuses on a particular aspect in the question. Success rate on each concept of initial submission is shown in Figure 3.2.2.1.

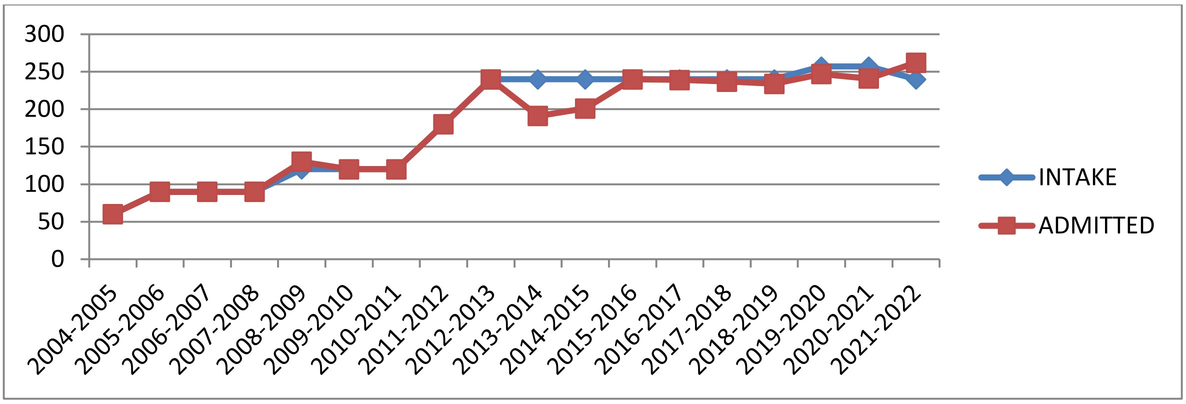
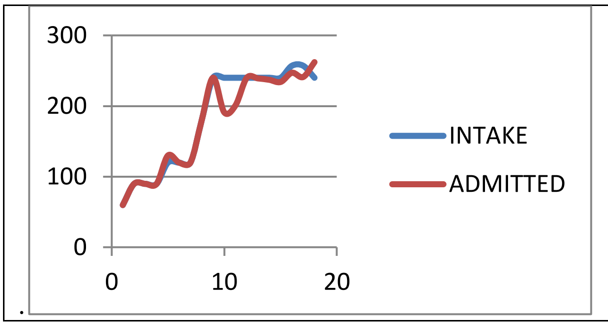


Figure 3.1.1 (b): Graph of intake and admitted data-sets for two decades



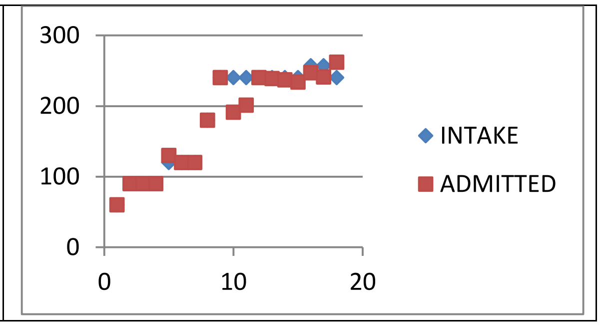


Figure 3.1.1 (c) **Figure 3.1.1 (d)**

Figure 3.1.1(c) and figure 3.1.1(d) shows the classification and pattern of distribution of the admission from 2004 to 2022, between 2016 to 2022 most of the data are concentrated and distributed at maximum value 240.

IV. CONCLUSION

ECE-CAPA provides instructors or course coordinators full access to the students' educational records. With this access, they are able to evaluate the problems presented in the course after the students have used the educational materials, through some statistical reports. ECE-CAPA also provides a quick review of students' submissions for every problem in a course. The instructor may monitor the number of submissions of every student in any homework set and its problems. The total numbers of solved problems in a homework set as compared with the total number of solved problems in a course are represented for every individual student.
ECE-CAPA reports a large volume of statistical information for every problem e.g., "total number of students who open the problem," "total number of submissions for the problem," "maximum number of submissions for the problem," "average number of submissions per problem," "number of students solving the problem correctly,".

ACKNOWLEDGEMENT

The authors are grateful to the "Aditya Educational Institutions and Cenrtal Christian University"

REFERENCES

[1]. Chernoff, H. (1973), "The use of Faces to represent Points in k-Dimensional Space Graphically", Journal of American Statistical Association, Vol. 68, page 361-368, June 1973.
[2]. Ching, J.Y. Wong, A.K.C. and Chan, C.C. (1995). "Class Dependent Discretisation for Inductive Learning from Continuous and Mixed-mode Data". IEEE Transaction. PAMI, 17(7) 641 - 645.
[3]. Dixon, J.K. (1979) "Pattern recognition with partly missing data" IEEE Transaction on Systems, Man and Cybernetics SMC 9, 617-621
[4]. Dhillon, I. S. and Modha, D. S., (2000)"A Data-clustering Algorithm on Distributed Memory Multiprocessors", In Proceedings of Large-scale Parallel KDD Systems Workshop, ACM SIGKDD, in Large-Scale Parallel Data Mining, Lecture Notes in Artificial Intelligence, 2000, 1759: 245-260.

[5]. Ester, M., Kriegel, H.-P., Xu. X. (1995) "A Database Interface for Clustering in Large Spatial Databases", Proceedings of the Knowledge Discovery and Data Mining Conference, pages 94-99, Montreal, Canada, 1995.
[6]. Goldberg, D.E. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning, MA, Addison-Wesley.
[7]. Lu, S. Y., and Fu, K. S. (1978). " A sentence-to-sentence clustering procedure for pattern analysis." IEEE Transactions on Systems, Man and Cybernetics SMC 8, 381-389.
[8]. Michalski, R.S., Stepp, R.E., III (1983), "Automated construction of classification: conceptual clustering versus numerical taxonomy" IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI 5, 396-410.
[9]. McQueen, J. B. (1967). "Some methods of classification and analysis of multivariate observations." Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281-297.
[10]. Clifton, Christopher "Encyclopedia Britannica: Definition of Data Mining" Retrieved 10 September 2012.
[11]. Application and Implementation of Data Mining in various field ,"Available at :http://en.wikipedia.org/wiki/Data_mining, 2012,(accessed on September 2012)
[12]. T. Friedman and M. Smith, "Measuring the Business Value of Data Quality," Gartner, no. G00218962, 2011.
[13]. L. Cai and Y. Zhu, "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era," Data Science Journal, vol. 14, no. 0, p. 2, 2015.
[14]. S. Chaudhuri, U. Dayal and V. Narasayya, "An overview of business intelligence technology," Communications of the ACM, vol. 54, no. 8, p. 88, 2011.
[15]. H. Chen, R. H. L. Chiang and V. C. Storey, "Business Intelligence and Analytics: From Big Data To Big Impact," Mis Quarterly, vol. 36, no. 4, pp. 1165-1188, 2012.
[16]. Interaction-Design, "Information Overload, Why it Matters and How to Combat It," 2017. [Online]. Available: https://www.interaction-design.org/literature/article/information-overloadwhy-it-matters-and-how-to-combat-it.
[17]. C. S. M. Batini, Data and Information Quality, Springer, 2016.
[18]. N. Askham, D. Cook, M. Doyle, H. Fereday, M. Gibson, U. Landbeck, R. Lee, C. Maynard, G. Palmer and J. Schwarzenbach, "The Six Primary Dimensions for Data Quality Assessment," 2013.
[19]. S. H. A. El-Sappagh, A. M. A. Hendawi and A. H. El Bastawissy, "A proposed model for data warehouse ETL processes," Journal of King Saud University - Computer and Information Sciences, vol. 23, no. 2, pp. 91-104, 2011.
[20]. C. Batini, C. Cappiello, C. Francalanci and A. Maurino, "Methodologies for data quality assessment and improvement," ACM Computing Surveys (CSUR), vol. 41, no. 3, p. 16:1–16:52, 2009.
[21]. C. Thota, G. Manogaran, D. Lopez and V. Vijayakumar, Big data security framework for distributed cloud data centers, 2016.
[22]. Asq, "Quality Glossary | ASQ," 2017. [Online]. Available: https://asq.org/qualityresources/quality-glossary.
[23]. McKinsey & Company, "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, no. June, p. 156, 2011.
[24]. M. J. Eppler and M. Helfert, "A classification and analysis of data quality costs," Quality, pp. 311- 325, 2004. [54] A. Haug, F. Zachariassen and D. van Liempd, The costs of poor data quality, vol. 4, 2011, pp. 168- 193.
[25]. N. V. Chawla, "Data Mining for Imbalanced Datasets: An Overview," in Data Mining and Knowledge Discovery Handbook, 2009, pp. 875-886.
[26]. M. M. Rahman and D. N. Davis, "Addressing the Class Imbalance Problem in Medical Datasets," International Journal of Machine Learning and Computing, pp. 224-228, 2013.
[27]. M. L. Bermingham, R. Pong-Wong, A. Spiliopoulou, C. Hayward, I. Rudan, H. Campbell, A. F. Wright, J. F. Wilson, F. Agakov, P. Navarro and C. S. Haley, "Application of high-dimensional feature selection: Evaluation for genomic prediction in man," Scientific Reports, vol. 5, 2015.
[28]. G. Tang, J. Pei, J. Bailey and G. Dong, "Mining multidimensional contextual outliers from categorical relational data," Intelligent Data Analysis, vol. 19, no. 5, pp. 1171-1192, 2015.
[29]. P. H. Shariat Panahy, F. Sidi, L. S. Affendey, M. A. Jabar, H. Ibrahim and A. Mustapha, "Discovering dependencies among data quality dimensions: A validation of instrument," Journal of Applied Sciences, vol. 13, no. 1, pp. 95-102, 2013.
[30]. P. H. Shariat Panahy, F. Sidi, L. S. Affendey, M. A. Jabar, H. Ibrahim and A. Mustapha, "A framework to construct data quality dimensions relationships," Indian Journal of Science and Technology, vol. 6, no. 5, pp. 4422-4431, 2013.
[31]. J. W. Graham, "Analysis of Missing Data," in Missing Data, 2012, pp. 978-1.
[32]. Adeniyi, D. A., Z. Wei, et al. (2014). "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method." Applied Computing and Informatics.
[33]. Aghakhani, N., A. S. Najar, et al. (2010). Supporting research collaboration through web based personal digital library. Computer Sciences and Convergence Information Technology (ICCIT), 2010 5th International Conference.
[34]. Ahn, H. J., H. Kang, et al. (2010). "Selecting a small number of products for effective user profiling in collaborative filtering." Expert Systems with Applications 37(4): 3055- 3062.
[35]. Akbar, M., C. A. Shaffer, et al. (2014). Recommendation based on deduced social networks in an educational digital library. Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries. London, United Kingdom, IEEE Press: 29-38.
[36]. Alam, S. (2011). Intelligent web usage clustering based recommender system. Proceedings of the fifth ACM conference on Recommender systems. Chicago, Illinois, USA, ACM: 367-370.
[37]. Amatriain, X., A. Jaimes, et al. (2011). Data Mining Methods for Recommender Systems. Recommender Systems Handbook. F. Ricci, L. Rokach, B. Shapira and P. B. Kantor. Boston, MA, Springer US: 39-71.
[38]. Ambayo, J. A. (2010). A Conceptual data mining model (DMM) used in Selective Dissemination of Information (SDI) : case study - Strathmore University Library. Faculty of Information Technology, Strathmore University, Kenya. MSc in Computer Based Information Systems: 50.
[39]. Anaraki, L. N. and A. Heidari (2011). Knowledge management process in digital age: Proposing a model for implementing e-learning through digital libraries. Application of Information and Communication Technologies (AICT), 2011 5th International Conference.
[40]. Arivazhagan and R. Pragaladan (2015). "Re-Adapted Apriori Algorithm in ECommerce Proposal Coordination." International Journal of Innovative Research in Computer and Communication Engineering 3(8): 7329-7336.
[41]. Barker, L. J. (2009). Science teachers' use of online resources and the digital library for Earth system education. Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries. Austin, TX, USA, ACM: 1-10.
[42]. Bidart, R., A. C. M. Pereira, et al. (2014). A Characterization of Access Profiles and Navigation in E-Commerce: A Tourism Application. Proceedings of the 20th Brazilian Symposium on Multimedia and the Web. João Pessoa, Brazil, ACM: 17-20.

[43]. Board, E. A. (2011). Redefining the Academic Library: Managing the Migration to Digital Information Services. Washington, DC, University Leadership Council.
[44]. Bobadilla, J., F. Ortega, et al. (2012). "A collaborative filtering approach to mitigate the new user cold start problem." Knowledge-Based Systems 26: 225-238.
[45]. Boddu, S. B., V. P. K. Anne, et al. (2010). Knowledge Discovery and Retrieval on World Wide Web Using Web Structure Mining. Mathematical/Analytical Modelling and Computer Simulation (AMS), 2010 Fourth Asia International Conference
[46]. Bonnin, G. and D. Jannach (2013). A Comparison of Playlist Generation Strategies for Music Recommendation and a New Baseline Scheme. Twenty-Seventh AAAI Conference on Artificial Intelligence. Washington USA, AAAI Publications.
[47]. Borhani-Fard, Z. M., Behrouz; Alinejad-Rokny, Hamid (2013). "Applying Clustering Approach in Blog Recommendation." Journal of Emerging Technologies in Web Intelligence 5(3). 147
[48]. Bundza, M. (2009). "Work of the Web Weavers: Web Development in Academic Libraries." Journal of Web Librarianship 3(3): 24.
[49]. Castro-Schez, J. J., R. Miguel, et al. (2011). "A highly adaptive recommender system based on fuzzy logic for B2C e-commerce portals." Expert Systems with Applications 38(3): 2441-2454.
[50]. Chen, Y., C. Wu, et al. (2011). "Solving the Sparsity Problem in Recommender Systems Using Association Retrieval." JOURNAL OF COMPUTERS 6(9): 1896-1902.
[51]. -Jen, W., C. Jen-Ming, et al. (2011). Using Web-Mining for Academic Measurement and Scholar Recommendation in Expert Finding System. Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on. 149
[52]. Choi, J. and L. Geehyuk (2009). New Techniques for Data Preprocessing Based on Usage Logs for Efficient Web User Profiling at Client Side. Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT '09. IEEE/WIC/ACM International Joint Conferences
[53]. Christidis, K. and G. Mentzas (2013). "A topic-based recommender system for electronic marketplace platforms." Expert Systems with Applications 40(11): 4370- 4379.
[54]. Crespo, R. G., O. S. Martínez, et al. (2011). "Recommendation System based on user interaction data applied to intelligent electronic books." Computers in Human Behavior 27(4): 1445-1449.

Dr Dola Sanjay S. “Data Representation, Features Extraction, Assessment and Prediction of ECE Admission in ACET." *International Journal of Engineering and Science*, vol. 13, no. 11, 2023, pp. 25-30.