REFERENCES

[1] Zhou, J., & Ye, J. M. (2020). Sentiment analysis in education research: a review of journal publications. Interactive learning environments, 1-13

[2] Arcinas, M. M., Sajja, G. S., Asif, S., Gour, S., Okoronkwo, E., & Naved, M. (2021). Role Of Data Mining In Education For Improving Students Performance For Social Change. Turkish Journal of Physiotherapy and Rehabilitation, 32(3), 6519-6526.

[3] de ANDRADE, T. L., Rigo, S. J., & Barbosa, J. L. V. (2021). Active Methodology, Educational Data Mining and Learning Analytics: A Systematic Mapping Study. Informatics in Education, 20(2), 171-203.

[4] Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. Telematics and Informatics, 37, 13-49.

[5] Rodrigues, M. W., Isotani, S., & Zarate, L. E. (2018). Educational Data Mining: A review of evaluation process in the e-learning. Telematics and Informatics, 35(6), 1701-1717.

[6][m10] Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. Education and information technologies, 26(1), 205-240.

[7] Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2020). Systematic ensemble model selection approach for educational data mining. Knowledge-Based Systems, 200, 105992.

[8] Hung, H. C., Liu, I. F., Liang, C. T., & Su, Y. S. (2020). Applying educational data mining to explore students' learning patterns in the flipped learning approach for coding education. Symmetry, 12(2), 213.

[9] Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2019). Implementing AutoML in educational data mining for prediction tasks. Applied Sciences, 10(1), 90.

[10] Kausar, S., Huahu, X., Hussain, I., Wenhao, Z., & Zahid, M. (2018). Integration of data mining clustering approach in the personalized E-learning system. IEEE Access, 6, 72724-72734.

[11] Buenaño-Fernandez, D., Villegas-CH, W., & Luján-Mora, S. (2019). The use of tools of data mining to decision making in engineering education—A systematic mapping study. Computer applications in engineering education, 27(3), 744-758.

[12] Feng, G., Fan, M., & Chen, Y. (2022). Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining. IEEE Access, 10, 19558-19571.

[13] Alshareef, F., Alhakami, H., Alsubait, T., & Baz, A. (2020). Educational data mining applications and techniques. International Journal of Advanced Computer Science and Applications, 11(4).

[14] Munk, M., & Drlík, M. (2016). Methodology of predictive modeling of students' behavior in virtual learning environment. In Formative Assessment, Learning Data Analytics and Gamification (pp. 187-216). Academic Press.

[15] Al Maani, D., & Shanti, Z. (2023). Technology-Enhanced Learning in Light of Bloom's Taxonomy: A Student-Experience Study of the History of Architecture Course. Sustainability, 15(3), 2624.

[16] Muhayimana, T., Kwizera, L., & Nyirahabimana, M. R. (2022). Using Bloom's taxonomy to evaluate the cognitive levels of Primary Leaving English Exam questions in Rwandan schools. Curriculum Perspectives, 42(1), 51-63.

[17] Ponto, H. (2022). The Effect of affective domain-based teaching on theoretical and practical outcomes of learning the basics of electricity at vocational high schools in North Sulawesi, Indonesia. International Journal on Integrated Education, 5(6), 8-20.

[18] Beena, B. R., & Suresh, E. S. M. (2022). Analysis of learning outcomes of Civil Engineering students of Kerala state using dimension reduction Techniques. Journal of Engineering Education Transformations, 35(Special Issue 1).

[19] Chiew, F. H., Noh, N., Oh, C. L., Noor, N. A. M., & Isa, C. M. M. (2022). Teaching, Learning and Assessments (TLA) in Civil Engineering Laboratory Courses in Open Distance Learning (ODL) during COVID-19 Pandemic. Asian Journal of University Education, 18(3), 818-829.

[20] Das, S., Das Mandal, S. K., & Basu, A. (2022). Classification of action verbs of Bloom's taxonomy cognitive domain: An empirical study. Journal of Education, 202(4), 554-566.

[21] Isa, C. M., Mustaffa, N. K., Joseph, E. O., & Preece, C. N. (2020). Development of Psychomotor Skill and Programme Outcome Attainment of Civil Engineering Students in Malaysia. Asian Journal of Vocational Education And Humanities, 1(2), 9-24.

[22] Serrat, O., & Serrat, O. (2017). Understanding and developing emotional intelligence. Knowledge solutions: Tools, methods, and approaches to drive organizational performance, 329-339.

[23] Behera, A. K. (2016). Understanding emotional intelligence in educational context. International Journal of Humanities and Social Science Invention, 5(2), 17-28.

[24] Thompson, E. R. (2020). Positive and negative affect schedule (PANAS). Encyclopedia of Personality and Individual Differences, 3963-3965.

[25] Monteiro, R. P., Coelho, G. L. D. H., Hanel, P. H., de Medeiros, E. D., & da Silva, P. D. G. (2022). The efficient assessment of self-esteem: proposing the brief rosenberg self-esteem scale. Applied Research in Quality of Life, 17(2), 931-947.

[26] Shogren, K. A., Little, T. D., Grandfield, E., Raley, S., Wehmeyer, M. L., Lang, K. M., & Shaw, L. A. (2020). The Self-Determination Inventory–Student Report: Confirming the factor structure of a new measure. Assessment for Effective Intervention, 45(2), 110-120.

[27] Eysenck, S. B., Barrett, P. T., & Saklofske, D. H. (2021). The junior eysenck personality questionnaire. Personality and Individual Differences, 169, 109974.

[28] Moeinaddini, M., Asadi-Shekari, Z., Aghaabbasi, M., Saadi, I., Shah, M. Z., & Cools, M. (2020). Proposing a new score to measure personal happiness by identifying the contributing factors. Measurement, 151, 107115.

[29] Zeng, G., Fung, S. F., Li, J., Hussain, N., & Yu, P. (2020). Evaluating the psychometric properties and factor structure of the general self-efficacy scale in China. Current Psychology, 1-11.

[20] El Bouchefry, K., & de Souza, R. S. (2020). Learning in big data: Introduction to machine learning. In Knowledge discovery in big data from astronomy and earth observation (pp. 225-249). Elsevier.

[31] Kong, L., Li, C., Ge, J., Zhang, F., Feng, Y., Li, Z., & Luo, B. (2020). Leveraging multiple features for document sentiment classification. Information Sciences, 518, 39-55.

[32] Wan, C., Peng, Y., Xiao, K., Liu, X., Jiang, T., & Liu, D. (2020). An association-constrained LDA model for joint extraction of product aspects and opinions. Information Sciences, 519, 243-259.

[33] Wang, X., Tang, M., Yang, T., & Wang, Z. (2021). A novel network with multiple attention mechanisms for aspect-level sentiment analysis. Knowledge-Based Systems, 227, 107196.

[34] Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of sentiment analysis in finance: from lexicons to transformers. IEEE access, 8, 131662-131682.

[35] Zhang, C., Tian, Y. X., Fan, Z. P., Liu, Y., & Fan, L. W. (2020). Product sales forecasting using macroeconomic indicators and online reviews: a method combining prospect theory and sentiment analysis. Soft Computing, 24(9), 6213-6226.

[36] Carvalho, J., & Plastino, A. (2021). On the evaluation and combination of state-of-the-art features in Twitter sentiment analysis. Artificial Intelligence Review, 54(3), 1887-1936.

[37] Majumder, N., Poria, S., Peng, H., Chhaya, N., Cambria, E., & Gelbukh, A. (2019). Sentiment and sarcasm classification with multitask learning. IEEE Intelligent Systems, 34(3), 38-43.

[38] Rajput, Q., Haider, S., & Ghani, S. (2016). Lexicon-based sentiment analysis of teachers' evaluation. Applied computational intelligence and soft computing, 2016.

[39] Nasim, Z., Rajput, Q., & Haider, S. (2017). Sentiment analysis of student feedback using machine learning and lexicon based approaches. In 2017 international conference on research and innovation in information systems (ICRIIS) (pp. 1-6). IEEE.

[40] Aung, K. Z., & Myo, N. N. (2017, May). Sentiment analysis of students' comment using lexicon based approach. In 2017 IEEE/ACIS 16th international conference on computer and information science (ICIS) (pp. 149-154). IEEE.

[41] Hong, J., Tamakloe, R., & Park, D. (2020). Application of association rules mining algorithm for hazardous materials transportation crashes on the expressway. Accident Analysis & Prevention, 142, 105497.

[42] Wang, T., Xiao, B., & Ma, W. (2022). Student Behavior Data Analysis Based on Association Rule Mining. International Journal of Computational Intelligence Systems, 15(1), 1-9.

[43] Czibula, G., Mihai, A., & Crivei, L. M. (2019). S PRAR: A novel relational association rule mining classification model applied for academic performance prediction. Procedia Computer Science, 159, 20-29.

[44] Zhang, Q., Lu, J., & Jin, Y. (2021). Artificial intelligence in recommender systems. Complex & Intelligent Systems, 7(1), 439-457.

[45] Pujahari, A., & Sisodia, D. S. (2022). Item feature refinement using matrix factorization and boosted learning based user profile generation for content-based recommender systems. Expert Systems with Applications, 206, 117849.

[46] Alhijawi, B., & Kilani, Y. (2020). A collaborative filtering recommender system using genetic algorithm. Information Processing & Management, 57(6), 102310.

[47] Dong, M., Zeng, X., Koehl, L., & Zhang, J. (2020). An interactive knowledge-based recommender system for fashion product design in the big data environment. Information Sciences, 540, 469-488.

[48] Seth, R., & Sharaff, A. (2022). A Comparative Overview of Hybrid Recommender Systems: Review, Challenges, and Prospects. Data Mining and Machine Learning Applications, 57-98.

[49] Bogarín, A., Cerezo, R., & Romero, C. (2018). A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(1), e1230.

[50] Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational data mining techniques for student performance prediction: method review and comparison analysis. Frontiers in psychology, 12, 698490.

[51] Czibula, G., Ciubotariu, G., Maier, M. I., & Lisei, H. (2022). IntelliDaM: A Machine Learning-Based Framework for Enhancing the Performance of Decision-Making Processes. A Case Study for Educational Data Mining. IEEE Access, 10, 80651-80666.

[52] Prada, M. A., Dominguez, M., Vicario, J. L., Alves, P. A. V., Barbu, M., Podpora, M., ... & Vilanova, R. (2020). Educational data mining for tutoring support in higher education: a web-based tool case study in engineering degrees. IEEE Access, 8, 212818-212836.

[53] Rahman, M. M., Watanobe, Y., Matsumoto, T., Kiran, R. U., & Nakamura, K. (2022). Educational data mining to support programming learning using problem-solving data. IEEE Access, 10, 26186-26202.

[54] Dabhade, P., Agarwal, R., Alameen, K. P., Fathima, A. T., Sridharan, R., & Gopakumar, G. (2021). Educational data mining for predicting students' academic performance using machine learning algorithms. Materials Today: Proceedings, 47, 5260-5267.

[55] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355.

[56] Riestra-González, M., del Puerto Paule-Ruíz, M., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. Computers & Education, 163, 104108.

[57] Karthikeyan, V. G., Thangaraj, P., & Karthik, S. (2020). Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation. Soft Computing, 24(24), 18477-18487.

[58] Crivei, L. M., Czibula, G., Ciubotariu, G., & Dindelegan, M. (2020, May). Unsupervised learning based mining of academic data sets for students' performance analysis. In 2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 000011-000016). IEEE.

[59] Delgado, S., Morán, F., San José, J. C., & Burgos, D. (2021). Analysis of Students' Behavior Through User Clustering in Online Learning Settings, Based on Self Organizing Maps Neural Networks. IEEE Access, 9, 132592-132608.

[60] Okoye, K., Arrona-Palacios, A., Camacho-Zuñiga, C., Achem, J. A. G., Escamilla, J., & Hosseini, S. (2022). Towards teaching analytics: a contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification. Education and Information Technologies, 1-43.

[61] Kumar, E. V., alias Balamurugan, S. A., & Sasikala, S. (2021). Multi-tier student performance evaluation model (MTSPEM) with integrated classification techniques for educational decision making. International Journal of Computational Intelligence Systems, 14(1), 1796-1808.

[62] Al Duhayyim, M., Marzouk, R., Al-Wesabi, F. N., Alrajhi, M., Hamza, M. A., & Zamani, A. S. (2022). An improved evolutionary algorithm for data mining and knowledge discovery. CMC-COMPUTERS MATERIALS & CONTINUA, 71(1), 1233-1247.

[63] Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1), 11.

[64] Sokkhey, P., & Okazaki, T. (2020). Developing web-based support systems for predicting poor-performing students using educational data mining techniques. International Journal of Advanced Computer Science and Applications, 11(7).

[65] Chui, K. T., Liu, R. W., Zhao, M., & De Pablos, P. O. (2020). Predicting students' performance with school and family tutoring using generative adversarial network-based deep support vector machine. IEEE Access, 8, 86745-86752.

[66] Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A. A., Abid, M., ... & Khan, S. U. (2021). Predicting at-risk students at different percentages of course length for early intervention using machine learning models. Ieee Access, 9, 7519-7539.

[67] Behr, A., Giese, M., Teguim K, H. D., & Theune, K. (2020). Early prediction of university dropouts–a random forest approach. Jahrbücher für Nationalökonomie und Statistik, 240(6), 743-789.

[68] Pek, R. Z., Özyer, S. T., Elhage, T., ÖZYER, T., & Alhajj, R. (2022). The Role of Machine Learning in Identifying Students At-Risk and Minimizing Failure. IEEE Access, 11, 1224-1243.

[69] Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. M. (2021). Multiclass prediction model for student grade prediction using machine learning. IEEE Access, 9, 95608-95621.

[70] Ko, C. Y., & Leu, F. Y. (2020). Examining successful attributes for undergraduate students by applying machine learning techniques. IEEE Transactions on Education, 64(1), 50-57.

[71] Rafique, A., Khan, M. S., Jamal, M. H., Tasadduq, M., Rustam, F., Lee, E., ... & Ashraf, I. (2021). Integrating learning analytics and collaborative learning for improving student's academic performance. IEEE Access, 9, 167812-167826.

[72] Lin, Y., Liu, H., Chen, Z., & Ma, K. (2020). Machine Learning-Based Classification of Academic Performance via Imaging Sensors. IEEE Sensors Journal, 21(22), 24952-24958.

[73] Alsariera, Y. A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., & Ali, N. A. (2022). Assessment and evaluation of different machine learning algorithms for predicting student performance. Computational Intelligence and Neuroscience, 2022.

[74] Kastrati, Z., Imran, A. S., & Kurti, A. (2020). Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs. IEEE Access, 8, 106799-106810.

[75] Tao, X., Shannon-Honson, A., Delaney, P., Li, L., Dann, C., Li, Y., & Xie, H. (2022). Data Analytics on Online Student Engagement Data for Academic Performance Modeling. Ieee Access, 10, 103176-103186.

[76] Zhai, G., Yang, Y., Wang, H., & Du, S. (2020). Multi-attention fusion modeling for sentiment analysis of educational big data. Big Data Mining and Analytics, 3(4), 311-319.

[77] Han, Z., Wu, J., Huang, C., Huang, Q., & Zhao, M. (2020). A review on sentiment discovery and analysis of educational big- data. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(1), e1328.

[78] Estrada, M. L. B., Cabada, R. Z., Bustillos, R. O., & Graff, M. (2020). Opinion mining and emotion recognition applied to learning environments. Expert Systems with Applications, 150, 113265.

[79] Spatiotis, N., Perikos, I., Mporas, I., & Paraskevas, M. (2020). Sentiment analysis of teachers using social information in educational platform environments. International Journal on Artificial Intelligence Tools, 29(02), 2040004.

[80] Asghar, M. Z., Ullah, I., Shamshirband, S., Khundi, F. M., & Habib, A. (2020). Fuzzy-based sentiment analysis system for analyzing student feedback and satisfaction. Computers, Materials & Continua, 62(2), 631-655.

[81] He, K., Mao, R., Gong, T., Li, C., & Cambria, E. (2022). Meta-based self-training and re-weighting for aspect-based sentiment analysis. IEEE Transactions on Affective Computing. (01), 1–13

[82] Nikolić, N., Grljević, O., & Kovačević, A. (2020). Aspect-based sentiment analysis of reviews in the domain of higher education. The Electronic Library, 38(1), 44-64.

[83] Melba Rosalind, J., & Suguna, S. (2022, July). Predicting Students' Satisfaction Towards Online Courses Using Aspect-Based Sentiment Analysis. In Computer, Communication, and Signal Processing: 6th IFIP TC 5 International Conference, ICCCSP 2022, Chennai, India, February 24–25, 2022, Revised Selected Papers (pp. 20-35). Cham: Springer International Publishing.

[84] OSMANOĞLU, U. Ö., Atak, O. N., Çağlar, K., Kayhan, H., & Talat, C. A. N. (2020). Sentiment analysis for distance education course materials: A machine learning approach. Journal of Educational Technology and Online Learning, 3(1), 31-48.

[85] Kaur, D. (2022). Incorporating sentimental analysis into development of a hybrid classification model: A comprehensive study. International Journal of Health Sciences, 6, 1709-1720.

[86] Sangeetha, K., & Prabha, D. (2021). Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM. Journal of Ambient Intelligence and Humanized Computing, 12, 4117-4126.

[87] Bilal, M., Ali, G., Iqbal, M. W., Anwar, M., Malik, M. S. A., & Kadir, R. A. (2022). Auto-Prep: Efficient and Automated Data Preprocessing Pipeline. IEEE Access, 10, 107764-107784.

[88] Alghamdi, T. A., & Javaid, N. (2022). A survey of preprocessing methods used for analysis of big data originated from smart grids. IEEE Access, 10, 29149-29171.

[89] Raja, P. S., & Thangavel, K. J. S. C. (2020). Missing value imputation using unsupervised machine learning techniques. Soft Computing, 24(6), 4361-4392.

[90] Bansal, P., Deshpande, P., & Sarawagi, S. (2021). Missing value imputation on multidimensional time series. arXiv preprint arXiv:2103.01600.

[91] McGinnis, W. D., Siu, C., Andre, S., & Huang, H. (2018). Category encoders: a scikit-learn-contrib package of transformers for encoding categorical data. Journal of Open Source Software, 3(21), 501.

[92] Duan, J. (2019). Financial system modeling using deep neural networks (DNNs) for effective risk assessment and prediction. Journal of the Franklin Institute, 356(8), 4716-4731.

[93] Munawar, H. S., Ullah, F., Qayyum, S., & Shahzad, D. (2022). Big data in construction: current applications and future opportunities. Big Data and Cognitive Computing, 6(1), 18.

[94] Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. J. Appl. Sci. Technol. Trends, 1(2), 56-70.

[95] Ayesha, S., Hanif, M. K., & Talib, R. (2020). Overview and comparative study of dimensionality reduction techniques for high dimensional data. Information Fusion, 59, 44-58.

[96] Espadoto, M., Martins, R. M., Kerren, A., Hirata, N. S., & Telea, A. C. (2019). Toward a quantitative survey of dimension reduction techniques. IEEE transactions on visualization and computer graphics, 27(3), 2153-2173.

[97] Huang, X., Wu, L., & Ye, Y. (2019). A review on dimensionality reduction techniques. International Journal of Pattern Recognition and Artificial Intelligence, 33(10), 1950017.

[98] Velliangiri, S., & Alagumuthukrishnan, S. J. P. C. S. (2019). A review of dimensionality reduction techniques for efficient computation. Procedia Computer Science, 165, 104-111.

[99] Jia, W., Sun, M., Lian, J., & Hou, S. (2022). Feature dimensionality reduction: a review. Complex & Intelligent Systems, 8(3), 2663-2693.

[100] Marbouti, F., Ulas, J., & Wang, C. H. (2020). Academic and demographic cluster analysis of engineering student success. IEEE Transactions on Education, 64(3), 261-266.

[101] Chang, W., Ji, X., Liu, Y., Xiao, Y., Chen, B., Liu, H., & Zhou, S. (2020). Analysis of university students' behavior based on a fusion K-means clustering algorithm. Applied Sciences, 10(18), 6566.

[102] Zhang, S., Shen, M., & Yu, Y. (2021). Research on student Big Data portrait method based on improved K-means algorithm. In 2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST) (pp. 146-150). IEEE.

[103] McBroom, J., Yacef, K., & Koprinska, I. (2020, June). DETECT: a hierarchical clustering algorithm for behavioral trends in temporal educational data. In Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part I (pp. 374-385). Cham: Springer International Publishing.

[104] Ahmed, A., Zualkernan, I., & Elghazaly, H. (2021, July). Unsupervised clustering of skills for an online learning platform. In 2021 International Conference on Advanced Learning Technologies (ICALT) (pp. 200-202). IEEE.

[105] Almasri, A., Alkhawaldeh, R. S., & Çelebi, E. (2020). Clustering-based EMT model for predicting student performance. Arabian Journal for Science and Engineering, 45, 10067-10078.

[106] Chaves, V. E. J., García-Torres, M., Alonso, D. B., Gómez-Vela, F., Divina, F., & Vázquez-Noguera, J. L. (2021). Analysis of student achievement scores via cluster analysis. In The 11th International Conference on EUropean Transnational Educational (ICEUTE 2020) 11 (pp. 399-408). Springer International Publishing.

[107] Chu, Y., & Yin, X. (2021). Data analysis of college students' mental health based on clustering analysis algorithm. Complexity, 2021, 1-10.

[108] Guo, H., & Wang, M. (2022). Analysis on the Penetration of Emotional Education in College Physical Education Based on Emotional Feature Clustering. Scientific Programming, 2022.

[109] Hassan, Y. M., Elkorany, A., & Wassif, K. (2022). Utilizing Social Clustering-Based Regression Model for Predicting Student's GPA. IEEE Access, 10, 48948-48963.

[110] Li, G., Alfred, R., & Wang, X. (2021). Student Behavior Analysis and Research Model Based on Clustering Technology. Mobile Information Systems, 2021, 1-6.

[11] Li, X., Zhang, Y., Cheng, H., Zhou, F., & Yin, B. (2021). An unsupervised ensemble clustering approach for the analysis of student behavioral patterns. Ieee Access, 9, 7076-7091.

[112] Priyambada, S. A., Er, M., Yahya, B. N., & Usagawa, T. (2021). Profile-based cluster evolution analysis: Identification of migration patterns for understanding student learning behavior. IEEE Access, 9, 101718-101728.

[113] Chi, D. (2021). Research on the Application of K-Means Clustering Algorithm in Student Achievement. In 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE) (pp. 435-438). IEEE.

[114] Lei, J. (2022). Association Rule Mining Algorithm in College Students' Quality Evaluation System. Journal of Electrical and Computer Engineering, 2022.

[115] Jia, N., & Madina, Z. (2022). An association rule-based multiresource mining method for MOOC teaching. Computational and Mathematical Methods in Medicine, 2022.

[116] Liu, H., & Chen, X. (2022). Construction and optimization of mental health education consultation management system based on decision tree association rule mining. Mathematical Problems in Engineering, 2022.

[117] Das, C., Bose, S., Chanda, A., Singh, S., Das, S., & Ghosh, K. (2021). Impact of prerequisite subjects on academic performance using association rule mining. In Progress in Advanced Computing and Intelligent Engineering: Proceedings of ICACIE 2019, Volume 2 (pp. 227-236). Springer Singapore.

[118] Salahli, M. A., Gasimzadeh, T., Alasgarova, F., & Guliyev, A. (2021). Analysis of Relationship Between Learning Outcomes and Student's Exam Results Using Association Rule Mining and Fuzzy Inference Rules. In 14th International Conference on Theory and Application of Fuzzy Systems and Soft Computing–ICAFS-2020 14 (pp. 354-361). Springer International Publishing.

[119] Shatnawi, R., Althebyan, Q., Ghaleb, B., & Al-Maolegi, M. (2021). A Student Advising System Using Association Rule Mining. International Journal of Web-Based Learning and Teaching Technologies (IJWLTT), 16(3), 65-78.

[120] [ch5-26] Rajab, K.D. (2019). New Associative Classification Method Based on Rule Pruning for Classification of Datasets. IEEE Access, 7, 157783–157795.

[121] A. Segatori, A. Bechini, P. Ducange and F. Marcelloni, "A Distributed Fuzzy Associative Classifier for Big Data," in IEEE Transactions on Cybernetics, vol. 48, no. 9, pp. 2656-2669, 2018

[122] S. Kumi, C. Lim, & S.G. Lee, "Malicious url detection based on associative classification", Entropy, vol. 23, no. 2, 2021

[123] S.P. Siddique Ibrahim, & M. Sivabalakrishnan, "An evolutionary memetic weighted associative classification algorithm for heart disease prediction", In Recent advances on memetic algorithms and its applications in image processing, Springer, pp. 183-199, 2020

[124] P.R. Pal, P. Pathak, & S. Luma-Osmani, "IHAC: Incorporating Heuristics for Efficient Rule Generation & Rule Selection in Associative Classification", Journal of Information & Knowledge Management, vol. 20, no. 1, 2021

[125] J. Mattiev, & B. Kavšek, "CMAC: clustering class association rules to form a compact and meaningful associative classifier", In International Conference on Machine Learning, Optimization, and Data Science, Springer, pp. 372-384, 2020

[126] Song, K.; Lee, K. (2017). Predictability-based collective class association rule mining. Expert Syst. Appl. 79, 1–7.

[127] Alwidian, J., Hammo, B.H., & Obeid, N. (2018). WCBA: Weighted classification based on association rules algorithm for breast cancer disease. Appl. Soft Comput. 62, 536–549

[128] Deschênes, M. (2020). Recommender systems to support learners' Agency in a Learning Context: a systematic review. International Journal of Educational Technology in Higher Education, 17(1), 50.

[129] Zhang, M., Liu, S., & Wang, Y. 2020. STR-SA: Session-based thread recommendation for online course forum with self-attention. In 2020 IEEE Global Engineering Education Conference (EDUCON) (pp. 374-381). IEEE.

[130] Mondal, B., Patra, O., Mishra, S., & Patra, P. (2020). A course recommendation system based on grades. In 2020 international conference on computer science, engineering and applications (ICCSEA) (pp. 1-5). IEEE.

[131] Cheng, Y., & Bu, X. (2020). Research on key technologies of personalized education resource recommendation system based on big data environment. In Journal of Physics: Conference Series (Vol. 1437, No. 1, p. 012024). IOP Publishing.

[132] Dhar, J., & Jodder, A. K. (2020). An Effective Recommendation System to Forecast the Best Educational Program Using Machine Learning Classification Algorithms. Ingénierie des Systèmes d Inf., 25(5), 559-568.

[133] Chen, H., Yin, C., Li, R., Rong, W., Xiong, Z., & David, B. (2019). Enhanced learning resource recommendation based on online learning style model. Tsinghua Science and Technology, 25(3), 348-356.

[134] Li, J., & Ye, Z. (2020). Course recommendations in online education based on collaborative filtering recommendation algorithm. Complexity, 2020, 1-10.

[135] Agbonifo, O. C., & Akinsete, M. (2020). Development of an ontology-based personalised E-learning recommender system. International Journal of Computer (IJC), 38(1), 102-112.

[136] Yanes, N., Mostafa, A. M., Ezz, M., & Almuayqil, S. N. (2020). A machine learning-based recommender system for improving students learning experiences. IEEE Access, 8, 201218-201235.

[137] Kang, Y., Cai, Z., Tan, C. W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. Journal of Management Analytics, 7(2), 139-172.

[138] PM, K. R. (2022). Sentiment analysis, opinion mining and topic modelling of epics and novels using machine learning techniques. Materials Today: Proceedings, 51, 576-584.

[139] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7), 5731-5780.

[140] Mirończuk, M. M., & Protasiewicz, J. (2018). A recent overview of the state-of-the-art elements of text classification. Expert Systems with Applications, 106, 36-54.

[141] Kastrati, Z., Dalipi, F., Imran, A. S., Pireva Nuci, K., & Wani, M. A. (2021). Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study. Applied Sciences, 11(9), 3986.

[142] Yang, D., Kraut, R. E., & Rose, C. P. (2016). Exploring the Effect of Student Confusion in Massive Open Online Courses. Journal of Educational Data Mining, 8(1), 52-83.

[143] Aziz, A. A., & Starkey, A. (2019). Predicting supervise machine learning performances for sentiment analysis using contextual-based approaches. IEEE Access, 8, 17722-17733.

[144] Poria, S., Majumder, N., Mihalcea, R., & Hovy, E. (2019). Emotion recognition in conversation: Research challenges, datasets, and recent advances. IEEE Access, 7, 100943-100953.

[145] Goleman, D. (1998). Working with emotional intelligence.Bantam, New York.

[146] Eysenck, S. B., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. Personality and individual differences, 6(1), 21-29.

[147] Hills, P., & Argyle, M. (2002). The Oxford Happiness Questionnaire: a compact scale for the measurement of psychological well-being. Personality and individual differences, 33(7), 1073-1082.

[148] Ji, J., Pang, W., Li, Z., He, F., Feng, G., & Zhao, X. (2020). Clustering mixed numeric and categorical data with cuckoo search. IEEE Access, 8, 30988-31003.

[149] Ramasamy, L. K., Kadry, S., Nam, Y., & Meqdad, M. N. (2021). Performance analysis of sentiments in Twitter dataset using SVM models. International Journal of Electrical and Computer Engineering (IJECE), 11(3), 2275-2284.

[150] Amala Jayanthi, M., & Shanthi I. (2020). Role of Educational data mining in student learning processes with sentiment analysis. International Journal of Knowledge and Systems Science, 11(4), 31–44

[151] Han, J., Pei, J., & Tong, H. (2022). Data mining: concepts and techniques. Morgan kaufmann.

[152] Jayanthi, M.A., Kumar, R. L., Surendran, A., & Prathap, K. (2016). Research contemplate on educational data mining. In 2016 IEEE International Conference on Advances in Computer Applications (ICACA) (pp. 110-114). IEEE.

[153] Dinh, D. T., Huynh, V. N., & Sriboonchitta, S. (2021). Clustering mixed numerical and categorical data with missing values. Information Sciences, 571, 418-442.

[154] Ramasamy, L. K., Kadry, S., & Lim, S. (2021). Selection of optimal hyper-parameter values of support vector machine for sentiment analysis tasks using nature-inspired optimization methods. Bulletin of Electrical Engineering and Informatics, 10(1), 290-298.

[155] Ahmad, A., & Khan, S. S. (2019). Survey of state-of-the-art mixed data clustering algorithms. Ieee Access, 7, 31883-31902.

[156] Brnawy, R., & Shiri, N. (2019). K-mixed prototypes: a clustering algorithm for relational data with mixed attribute types. In Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing (pp. 542-545).

[157] Huang, Z. (1997). Clustering large data sets with mixed numeric and categorical values. In Proceedings of the 1st pacific-asia conference on knowledge discovery and data mining,(PAKDD) (pp. 21-34).

[158] Ji, J., Bai, T., Zhou, C., Ma, C., & Wang, Z. (2013). An improved k-prototypes clustering algorithm for mixed numeric and categorical data. Neurocomputing, 120, 590-596.

[159] Chen, J. Y., & He, H. H. (2016). A fast density-based data stream clustering algorithm with cluster centers self-determined for mixed data. Information Sciences, 345, 271-293.

[160] Manochandar, S., Punniyamoorthy, M., & Jeyachitra, R. K. (2020). Development of new seed with modified validity measures for k-means clustering. Computers & Industrial Engineering, 141, 106290.

[161] Movahedi, M., Homayounfar, M., Fadaei, M., & Soufi, M. (2022). Developing a hybrid model for comparative analysis of financial data clustering algorithms. Journal of Decisions and Operations Research.

[162] Van de Velden, M., Iodice D'Enza, A., & Markos, A. (2019). Distance- based clustering of mixed data. Wiley Interdisciplinary Reviews: Computational Statistics, 11(3), e1456.

[163] Que, X., Jiang, S., Yang, J., & An, N. (2021). A similarity measurement with entropy-based weighting for clustering mixed numerical and categorical datasets. Algorithms, 14(6), 184.

[164] Li, X., Wu, Z., Zhao, Z., Ding, F., & He, D. (2021). A mixed data clustering algorithm with noise-filtered distribution centroid and iterative weight adjustment strategy. Information Sciences, 577, 697-721.

[165] Jia, H., & Cheung, Y. M. (2017). Subspace clustering of categorical and numerical data with an unknown number of clusters. IEEE transactions on neural networks and learning systems, 29(8), 3308-3325.

[166] D'urso, P., & Massari, R. (2019). Fuzzy clustering of mixed data. Information Sciences, 505, 513-534.

[167] Rodriguez- Sanchez, F., Bielza, C., & Larrañaga, P. (2022). Multipartition clustering of mixed data with Bayesian networks. International Journal of Intelligent Systems, 37(3), 2188-2218.

[168] Ryu, H. C., & Jung, S. (2021). MCF Tree-Based Clustering Method for Very Large Mixed-Type Data Set. IEEE Access, 9, 138580-138597.

[169] Ji, J., Li, R., Pang, W., He, F., Feng, G., & Zhao, X. (2021). A multi-view clustering algorithm for mixed numeric and categorical data. IEEE Access, 9, 24913-24924.

[170] Lee, Y., Park, C., & Kang, S. (2022). Deep Embedded Clustering Framework for Mixed Data. IEEE Access, 11.

[171] Ji, J., Chen, Y., Feng, G., Zhao, X., & He, F. (2019). Clustering mixed numeric and categorical data with artificial bee colony strategy. Journal of Intelligent & Fuzzy Systems, 36(2), 1521-1530.

[172] Rehman, S., & Sharma, A. (2017). Privacy Preserving Data Mining Using Association Rule Based on Apriori Algorithm. In Advanced Informatics for Computing Research, 712, 218–226.

[173] Liu, L., Yu, S., Wei, X., & Ning, Z. (2018). An improved Apriori–based algorithm for friends recommendation in microblog. International Journal of Communication Systems, 31(2), e3453.

[174] Thurachon, W., & Kreesuradej, W. (2021). Incremental association rule mining with a fast incremental updating frequent pattern growth algorithm. IEEE Access, 9, 55726-55741.

[175] Jijo, B. T., & Abdulazeez, A. M. (2021). Classification based on decision tree algorithm for machine learning. evaluation, 6, 7.

[176] Kumar, M. S., Dhulipala, V. S., & Baskar, S. (2021). Fuzzy unordered rule induction algorithm based classification for reliable communication using wearable computing devices in healthcare. Journal of Ambient Intelligence and Humanized Computing, 12, 3515-3526.

[177] Padillo, F., Luna, J. M., & Ventura, S. (2019). Evaluating associative classification algorithms for Big Data. Big Data Analytics, 4, 1-27.

[178] Hadi, W. E., Issa, G., & Ishtaiwi, A. (2017). ACPRISM: Associative classification based on PRISM algorithm. Information Sciences, 417, 287-300.

[179] Cao, W., Zhong, Q., Li, H., & Liang, S. (2020). A novel approach for associative classification based on information entropy of frequent attribute set. IEEE Access, 8, 140181-140193.

[180] Abdelhamid, N., Jabbar, A. A., & Thabtah, F. (2016, August). Associative classification common research challenges. In 2016 45th International Conference on Parallel Processing Workshops (ICPPW) (pp. 432-437). IEEE.

[181] Telikani, A., Gandomi, A. H., & Shahbahrami, A. (2020). A survey of evolutionary computation for association rule mining. Information Sciences, 524, 318-352.

[182] Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD international conference on Management of data (pp. 207-216).

[183] Diaz-Garcia, J. A., Ruiz, M. D., & Martin-Bautista, M. J. (2020). Non-query-based pattern mining and sentiment analysis for massive microblogging online texts. IEEE Access, 8, 78166-78182.

[184] Yu, L.A., Zhang, Y.D.: Weight-selected attribute bagging based on association rules for credit dataset classifcation. Syst. Eng. Theory Pract. 40(2), 366–372.

[185] Cao, L., Xu, L., Yang, F., Jia, P.F.: (2021) Infuencing factors analysis of pavement damage based on mining association rules. Comput. Syst. Appl. 30(1), 186-193

[186] Li, X. (2018) Applied research on strong association rules in personalized information push service of smart library. Inf. Sci. 36(4), 95–99.

[187] Ariannezhad, A., Wu, Y.J. (2020) Large-scale loop detector troubleshooting using clustering and association rule mining. J. Transp. Eng. A Syst. 146(7), 04020064.

[188] Guo, B., Li, Z.M., Zhang, J., Yu, Z.W. (2020) Cross-modal crowd sourced data for context-based scenic route recommendation. J. Zhengzhou Univ. (Nat. Sci. Ed.) 52(2), 22–28.

[189] Liu, B., Yiming, M., & Hsu, W. (1998). Integrating Classification and Association Rule Mining. In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 27–31

[190] Li, W., Han, J., & Pei, J. (2001). CMAR: Accurate and efficient classification based on multiple class-association rules. In Proceedings of the 2001 IEEE International Conference on Data Mining, San Jose, CA, USA, 29 November–2 December 2001; pp. 369–376.

[191] Abdelhamid, N. (2015). Multi-label rules for phishing classification. Appl. Comput. Inform. 11, 29–46.

[192] Hadi, W., Aburub, F., Alhawari, S. (2016). A new fast associative classification algorithm for detecting phishing websites. Appl. Soft Comput. 48, 729–734

[193] Thanajiranthorn, C., & Songram, P. (2020). Efficient rule generation for associative classification. Algorithms, 13(11), 299.

[194] Neysiani, B.S., Soltani, N., Mofidi, R., & Nadimi-Shahraki, M.H. "Improve performance of association rule-based collaborative filtering recommendation systems using genetic algorithm", International Journal of Information Technology and Computer Science, vol. 11, no. 2, pp. 48-55, 2019

[195] Varzaneh, H.H., Neysiani, B.S., Ziafat, H., & Soltani, N. "Recommendation systems based on association rule mining for a target object by evolutionary algorithms", Emerging Science Journal, vol. 2, no. 2, pp. 100-107, 2018

[196] Lops, P., Jannach, D., Musto, C., Bogers, T., & Koolen, M. "Trends in content-based recommendation", User Modeling and User-Adapted Interaction, vol. 29, no. 2, pp. 239-249, 2019

[197] Khanal, S.S., Prasad, P.W.C., Alsadoon, A., & Maag, A. "A systematic review: machine learning based recommendation systems for e-learning", Education and Information Technologies, vol. 25, no. 4, pp. 2635-2664, 2020

[198] Afolabi, I.T., Ayo, A., & Odetunmibi, O.A. "Academic Collaboration Recommendation for Computer Science Researchers Using Social Network Analysis", Wireless Personal Communications, vol. 121, no. 1, pp. 487-501, 2021

[199] Al Fararni, K., Nafis, F., Aghoutane, B., Yahyaouy, A., Riffi, J., & Sabri, A. "Hybrid recommender system for tourism based on big data and AI: A conceptual framework", Big Data Mining and Analytics, vol. 4, no. 1, pp. 47-55, 2021

[200] Oliveira, A., Teixeira, M. and Neto, C., "Recommendation of Educational Content to Improve Student Performance: An Approach based on Learning Styles", In Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020) -vol. 2, pp. 359-365, 2020

[201] Lu, J., Zhang, Q., & Zhang, G. "Recommender Systems: Advanced Developments", World Scientific Publishing, 2021

APPENDIX - 1

Sample Code – SentimentAnalyzer

```
package servlet1;
import java.util.Properties;
import org.ejml.simple.SimpleMatrix;
import edu.stanford.nlp.ling.CoreAnnotations;
import edu.stanford.nlp.neural.rnn.RNNCoreAnnotations;
import edu.stanford.nlp.pipeline.Annotation;
import edu.stanford.nlp.pipeline.StanfordCoreNLP;
import edu.stanford.nlp.sentiment.SentimentCoreAnnotations;
import edu.stanford.nlp.trees.Tree;
import edu.stanford.nlp.util.CoreMap;
/**
*
* @author jayanthi
*/
public class SentimentAnalyzer
{
static Properties props;
static StanfordCoreNLP pipeline;
public void initialize(String path)
{
// creates a StanfordCoreNLP object, with POS tagging, lemmatization, NER,
parsing, and sentiment
props = new Properties();
props.setProperty("parse.model",
path+"edu\\stanford\\nlp\\models\\lexparser\\englishPCFG.ser.gz");
props.setProperty("sentiment.model",
path+"edu\\stanford\\nlp\\models\\sentiment\\sentiment.ser.gz");
props.setProperty("annotators", "tokenize, ssplit, parse, sentiment");
pipeline=new StanfordCoreNLP(props);
//LexicalizedParserlp =
LexicalizedParser.loadModel("edu/stanford/nlp/models/lexparser/englishPCFG.ser.g
z");
}
public SentimentResultgetSentimentResult(String text) {
SentimentResultsentimentResult = new SentimentResult();
SentimentClassificationsentimentClass = new SentimentClassification();
if (text != null &&text.length() > 0) {
// run all Annotators on the text
Annotation annotation = pipeline.process(text);
for (CoreMap sentence:
annotation.get(CoreAnnotations.SentencesAnnotation.class)) {
// this is the parse tree of the current sentence
Tree tree = sentence.get(SentimentCoreAnnotations.SentimentAnnotatedTree.class);
SimpleMatrixsm = RNNCoreAnnotations.getPredictions(tree);
```

```
String sentimentType =
sentence.get(SentimentCoreAnnotations.SentimentClass.class);
sentimentClass.setVeryPositive((double)Math.round(sm.get(4)*100d));
sentimentClass.setPositive((double)Math.round(sm.get(3)*100d));
sentimentClass.setNeutral((double)Math.round(sm.get(2)*100d));
sentimentClass.setNegative((double)Math.round(sm.get(1)*100d));
sentimentClass.setVeryNegative((double)Math.round(sm.get(0)*100d));
sentimentResult.setSentimentScore(RNNCoreAnnotations.getPredictedClass(tree));
sentimentResult.setSentimentType(sentimentType);
sentimentResult.setSentimentClass(sentimentClass);
}}
Return sentimentResult;
}}
```

APPENDIX – 2

ETHICAL COMMITTEE APPROVAL

INSTITUTIONAL HUMAN ETHICS COMMITTEE

Avinashilingam

Institute for Home Science and Higher Education for Women
(Deemed to be university under Category 'A' by MHRD, Estd. u/s 3 of UGC Act 1956) Re-accredited with 'A^{++}' Grade by NAAC.
Recognised by UGC Under Section 12 B
Coimbatore- 641043, Tamil Nadu, India

Chairman
Dr.Sudha Ramalingam
Director – Research and Innovation
Professor- Community Medicine,
PSG Institute of Medical Sciences
& Research, Coimbatore

Member Secretary
Dr. A Thirumani Devi
Professor
Department of Food Science and Nutrition

Members
Mr.K. Arulmoli (Legal Expert)
Dr. Subashini K Sripathi
Dr. A Saraswathy(Medical Officer)
Ms. D. Kavitha
Dr. A R Sudamani Ramasamy
Dr. G. Victoria Naomi
Dr. Judith Justin
Dr. Anitha Subash
Dr.K. Sampath Rani

18.05.2023

To
Ms. M. Amala Jayanthi
Department of Computer Science
Avinashilingam Institute for Home Science and
Higher Education for Women
Coimbatore- 641043

Dear M. Amala Jayanthi

Ref: Your proposal No. IHEC/22-23/CS-06 entitled "An Integrated Framework for Students academic Performance Prediction Based on Affective and Cognitive State Using Mixed Numeric and Categorical Data" submitted for approval of IHEC on 08.04.2023.

The Institutional Human Ethics Committee of our University hereby grants approval to your research proposal No. IHEC/22-23/CS-06 entitled "An Integrated Framework for Students academic Performance Prediction Based on Affective and Cognitive State Using Mixed Numeric and Categorical Data". The Approval number for the same is AUW/IHEC/CS-22-23/XPD-06.

We wish you all the best in your research endeavours.

Regards

18/5/23

Dr. A Thirumani Devi
Member Secretary

APPENDIX - 3

PSYCOLOGY EXPERT LETTER

From
Dr. S. John Michael Raj,
Supervisor & Guide,
Professor (Retd.),
Dept. of Psychology,
Bharathiar University, Coimbatore – 641 046,
Residence: 29, BK Chetty Street, Fort,
Coimbatore-641 001.

15..06..2022

To
Members of Doctoral Committee,
Candidate Name: Ms. M. Amala Jayanthi,
Avinashilingam Deemed to be University for Women,
Coimbatore-641 043.

Sir/Madam,

Sub: The Psychological domains in the PhD work– Regarding
Ref: Discussion held on 15.06.2022.

With reference to the discussion had with the PhD Scholar Ms. M. Amala Jayanthi regarding the usage of Psychological Paradigms in her research plan I am to state the following points:

- The proposal proposed in this research is very clear and candid.
- The hypotheses generated are pertinent to the focus of the study.
- The selected Psychological Instruments in Questionnaire format are standardized ones to bring out the multiphase nature of the profiles of the subjects of the study.
- The statistical techniques selected and used to analyze the collected data are apt and appropriated.
- The study would really help in understanding the emotional, behavioral and personality attributes of the subjects studied.

With regards,

Yours faithfully,

(DR.S.JOHN MICHAEL RAJ)

Dr. S. JOHN MICHAEL RAJ, Ph.D.,
Former Professor,
Dept. of Psychology,
Bharathiar University,
Coimbatore - 641 046.

LIST OF PUBLICATIONS

- Amala Jayanthi, M., & Shanthi, I. (2022). Role of educational data mining in student learning processes with sentiment analysis: a survey. In Research Anthology on Interventions in Student Behavior and Misconduct (pp. 412-427). IGI Global.
- Jayanthi M, A., & Shanthi I, E. (2022). Quest_SA: Preprocessing Method for Closed-Ended Questionnaires Using Sentiment Analysis through Polarity. Mobile Information Systems, 2022.
- Amala Jayanthi, M., & Shanthi, I. (2023). Reclust: an efficient clustering algorithm for mixed data based on reclustering and cluster validation. Indonesian Journal of Electrical Engineering and Computer Science (pp. 545~552)

LIST OF CONFERENCES ATTENDED AND PAPERS PRESENTED:

1. Impact of Personality Traits on Students 'Academic Performance , presented in the International Conference on High Performance and Intelligent Computing organized by PSG college of Technology on Dec 9 2022.
2. A Study on the Impact of Self-Determination on Academic Achievement , presented in International Conference on Artificial Intelligence & Data Analytics, Internet of Things, Cyber Security (ICAIC'22)organized by Kumaraguru College of Technology, Coimbatore on May 21 & 22, 2022.
3. 3.ARM-pred- an efficient prediction based on rule generation for clustered data through associative classifier in the National Conference on Multidisciplinary Research for Sustainable Innovations (NCMRSI 2023) conducted by the Department of Computer Science, on 10th February 2023
4. Recommender System For Predicting Student Personality With Emotion Based On Closed-Ended Questionnaire Using Sentiment Analysis And Association Rule Mining in IEEE Conference on Application of AI insustainable computing- ITT2023 , Higher Colleges of Technology, United Arab Emirates.

Avinashilingam Institute for HomeScience and Higher Education for Women
(Deemed to be University Estd. u/s 3 of UGC Act 1956, Category 'A' by MHRD
Re-accredited with A++ Grade by NAAC. CGPA 3.65/4, Category I by UGC
Coimbatore - 641 043, Tamil Nadu, India

Appendix L2

(Item No 5 of Check List) Details of Research Publications

S.No	Article	Journal	Other Details Vol/No/Page No/ Year	Published in UGC-CARE / Scopus Indexed/ Web of Science
1	Role of educational datamining in Student Learning process with Sentiment Analysis A Survey	International Journal of knowledge & System Science	Volume 11 Issue 4 Oct-Dec 2020	Scopus
2.	Quest_SA: preprocessing Method for closed ended Questionnaires Using Sentiment analysis through Polarity	Hindawi Mobile Information Systems	ID:4733550 Volume:2022 Sep 2022	Scopus
3.	Reclust: An efficient Clustering algorithm for mixed data based on reclustering and cluster validation.	Indonesian Journal of Electrical, Engineering and Computer Science	Volume:29 No:1 Jan 2023 PP:545-552 ISSN 2502-4752	Scopus

*Proof of list of Journals from Internet to be attached along with copies of reprints.

Scholar : M. Amala Jayanthi
17/3/2023

Supervisor :
17/3/2023

Checked By:
20/3/2023
HoD/Dean of Respective School

The scholar Amala Jayanthi, M published her article in International Journal of knowledge and System science in Oct 2020 and active in Scopus from 2017 - ongoing

P.T.O.

② In mobile Information Systems published in september 2022 and active in scopus from 2005 - ongoing

③ In Indonesian Journal of Electrical Engineering and comp science in January 2023 and active in scopus from 2015 - ongoing

17.3.23

V. SELVANAYAGI

ASST. LIBRARIAN (SS)

International Journal of Knowledge and Systems Science
Volume 11 • Issue 4 • October-December 2020

Role of Educational Data Mining in Student Learning Processes With Sentiment Analysis:
A Survey

Amala Jayanthi M., Kumaraguru College of Technology, India

Elizabeth Shanthi I., Avinashilingam Institution for Home Science and Higher Education for Women, Avinashilingam University, India

ABSTRACT

Educational data mining is a research field that is used to enhance education system. Research studies using educational data mining are in increase because of the knowledge acquired for decision making to enhance the education process by the information retrieved by machine learning processes. Sentiment analysis is one of the most involved research fields of data mining in natural language processing, web mining, and text mining. It plays a vital role in many areas such as management sciences and social sciences, including education. In education, investigating students' opinions, emotions using techniques of sentiment analysis can understand the students' feelings that students experience in academic, personal, and societal environments. This investigation with sentiment analysis helps the academicians and other stakeholders to understand their motive on education is online. This article intends to explore different theories on education, students' learning process, and to study different approaches of sentiment analysis academics.

KEYWORDS

Education, Learning Theories, Sentiment Analysis, Student Emotions

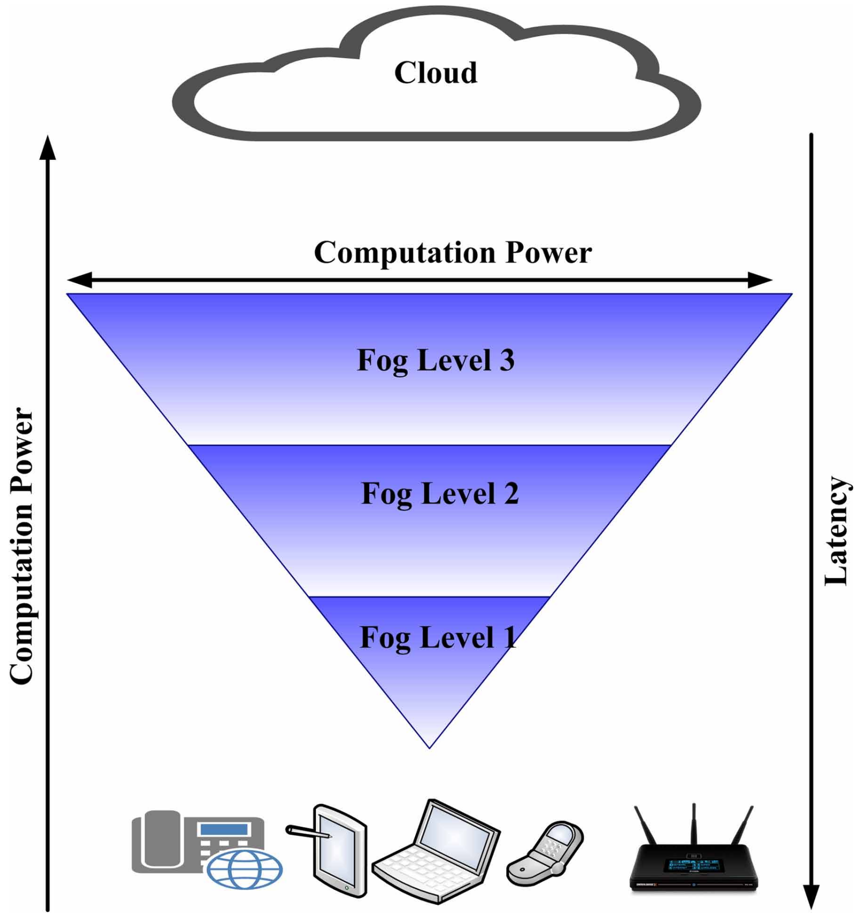
1. INTRODUCTION

Education is the continuous process of learning or ongoing development in abilities, values, attitudes, behaviours and intelligence. Types of Education include instruction, reading, narrative, debate, and guided study. Education is mostly provided under the inspection of teachers, and also but learners may undergo self-learning. The process of education can be done in a formal or informal environment. The definition of education differs between philosophers. Figure 1 shows some description of Education by various philosophers.

The essential motive of any academic program is to provide students with the required knowledge and skillsets to transform onto a productive professional within a given stipulated period. Using Data Mining (DM) strategies to evaluate knowledge about students may help establish potential explanations

DOI: 10.4018/IJKSS.2020100103

Figure 1. Definition of education



for learning. Educational Data Mining (EDM) is a data mining technology field developed to address educational problems (Altrabsheh et al., 2013). This contains a large volume of contextual information which offers a better understanding of learners based on their methods of learning. This utilizes DM methods to analyze data from the academic environment and to address academic problems that stop from achieving the motive of the educational programme. As with other extraction methods using DM techniques, EDM derives relevant, interpretable, valuable and novel information from educational data (Algarni 2016). Educational data mining researchers view the following as their work goals (T. E. D. Mining 2012):

- Predicting the potential learning behaviour of students by developing models of students that include such specific details as the experience, motivation, meta- cognition, and attitudes of students
- The discovery or development of domain structures characterizing the learning material and optimal instructional sequences.
- Investigating the influences of different kinds of pedagogical aid that learning applications can offer.
- Advancing empirical information about learning and learners by developing computational models that include student simulations, the environment and the pedagogy of the program

Sentiment analysis, also termed as opinion mining, is an application of natural language processing, computational linguistics, and content interpretation that, by analyzing the viewpoint, recognizes and retrieves emotion polarity from the content. The polarity of opinion is typically either positive (confident and enthusiastic) or negative (confused, boisterous, and furious), but often used as neutral (Altrabsheh et al., 2014). Mostly the sentiment analysis is applied in e-commerce, to analyze the customer reviews. Only a few articles (Altrabsheh et al., 2013; Altrabsheh et al., 2014) apply sentiment analysis in the education domain. This paper reviews sentiment analysis in the field of education.

The study is organized in this survey paper as the following. Section 2 gives the different definition of education Section 3 describes various theories on learning. It is based on emotion and academic performance of the student. Section 4 explains the research on Education, which mainly includes academic performance, student behaviour and emotions. The application of Education under computer science is described in Section 4. In section 5 the sentiment analysis and research in education using sentiment analysis is explored, and finally, section 5 concludes this paper.

2. WHAT IS EDUCATION?

Oxford definition of Education, Education is the process of acquiring or providing systematic instruction, especially at an academic environment like school, college or university. Education is the process of delivery of knowledge, skills and information to students by teachers.

According to Johann Heinrich Pestalozzi, a Swiss pedagogue and educational reformer, “Education is the natural, harmonious and progressive development of man’s innate powers.”

John Dewey, an American educational reformer who was also a philosopher and psychologist whose thoughts have been instrumental in Education and social reform. He stated, “Education is the development of all those capacities in the individual, which will enable him to control his environment and fulfil his possibilities”.

Based on John Adams, was an American statesman, attorney, diplomat, writer, and the second president of the United States, “Education is a conscious and deliberate process in which one personality acts upon another to modify the development of that other by the communication and manipulation of knowledge.”

3. THEORIES ON LEARNING

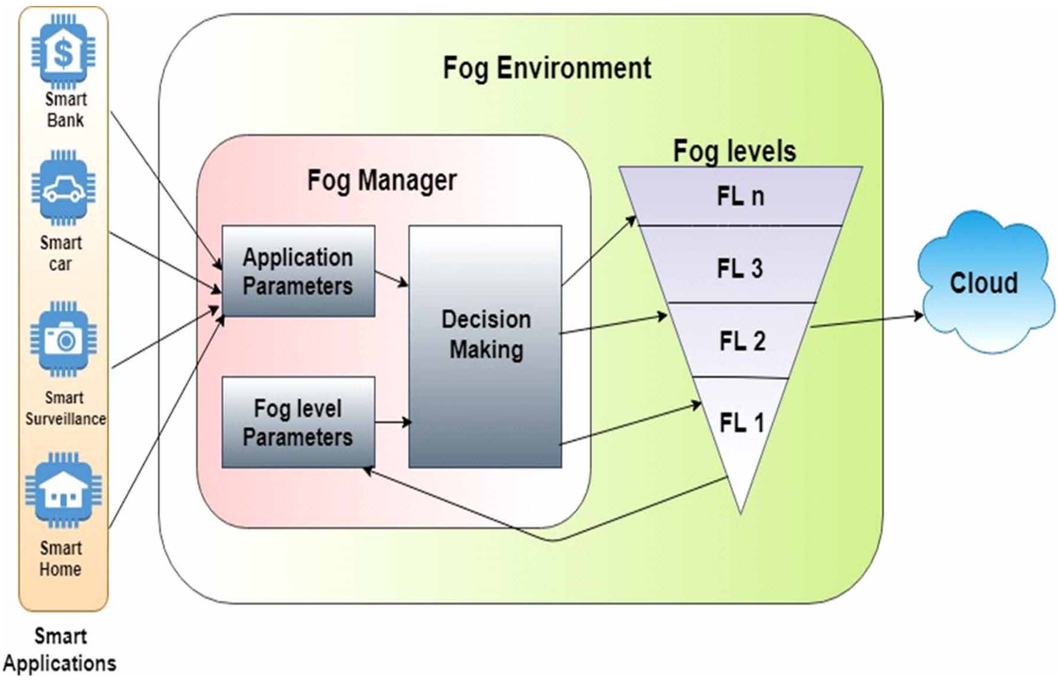
The philosophy of Education is the principle of intention, practice and perception of teaching and learning. Theory of learning explains the process of consuming, processing and maintaining information while learning. The factors of Cognitive, emotional and environmental and prior experience play a role in how opinion or a world view is gained or changed and information and skills conserved.

Theories of learning tend to fall under one of a variety of viewpoints or paradigms, including behaviourism, cognitivism, constructivism, humanism, connectivism, concise and others. This section discusses three fundamental theories shown in Figure 2 and also explains some taxonomy of learning theories.

Behaviourism is a perspective of which contextual influences can describe behaviour, and behavioural conditioning should be seen as a primary learning mechanism. The principles of positive and negative feedback in behaviourism are essential methods for understanding and altering behaviour, as well as a method of discipline and reward. It involves repetitive actions, overt encouragement and participatory rewards. This is excellent for developing guidelines, especially for managing behaviour.

Cognitivism is a theory of learning by Jean Piaget. In this, he states that the child builds detailed neural processes and behavioural response to experiences. Students learn most efficiently in this theory

Figure 2. Learning theories



by reading text and through teaching in lecture. In the opinion of cognitivism, learning comes about when the student reorganizes knowledge, either by seeking new approaches or by modifying old ones.

Constructivism believes that individuals are responsible for expanding their self-view on the environment by applying what they learned in the course of relating current knowledge to other experiences founded on previous experiences. People make use of those intercommunications and new knowledge to form their interpretation. Since students build their knowledge base, it is not always possible to predict outcomes, so the instructor will test and question the myths that may have arisen. A constructivistic approach may not be the best method to use where reliable outcomes are expected. Table 1 shows a short comparison of these three learning theories.

3.1 Bloom's Taxonomy of Learning Domain

In 1956, Bloom's taxonomy was by developed by Dr Benjamin Bloom (Bloom n.d.) educational psychologist to promote higher modes concerning educational thought, such as evaluating and analysing, rather than merely memorising facts. The three types of knowledge, or domains, they describe are cognitive (knowledge), affective (attitudes), and psychomotor (physical abilities).

Cognitive Domain involves awareness, and analytical skills growth. This includes remembering or understanding basic facts, organizational habits and principles that help to improve analytical skills and competencies. The affective domain is concerned with emotions such as feelings, values, appreciation, enthusiasm, motivations and attitudes. The psychomotor environment applies to all aims related to the interpretive gestures and discrete physical processes of reflex actions (Figure 3).

3.2 Gagné's Conditions of Learning

The book entitled "Robert Mills Gagné published the Conditions of Learning was an American educational psychologist in 1965. He has discussed the analysis of learning objectives and how the different classes of objective require specific teaching methods in his book.

Table 1. Learning theories comparison

	Behaviourism	Cognitivism	Constructivism
Knowledge	Knowledge is a collection of behavioural responses to stimuli in the world	Learners are deliberately developing information systems with cognitive structures, built on pre-existing cognitive frameworks.	Knowledge is developed inside social contexts by experiences with a community of Knowledge
Learning	The Learner unconsciously consumes a predefined body of knowledge. Fostered by repetition and positive reinforcement	Effective assimilation of new knowledge and integration of existing cognitive systems. The learners stress discovery.	Integrating the students into a community of knowledge. Collaborative assimilation and absorption of new knowledge
Motivation	Extrinsic feedback, involving both positive and negative.	Intrinsic; the learners have their own goals and are motivated to understand.	Intrinsic and Extrinsic. Learning goals and motivations are decided by the intelligence group, both by learners and by extrinsic incentives.
Suggestion	The instructor conveys right interpersonal responses and is learned by the students.	The teacher makes learning possible by creating an atmosphere that facilitates experimentation and assimilation/ accommodation	The Instructor encourages and directs collective instruction. Encourage social practice.

He proposes five conditions of learning, which fall under the cognitive, affective and psycho-motor domains discussed by Bloom.

Gagné's Conditions of Learning are as follows:

- Verbal information (Cognitive domain)
- Intellectual skills (Cognitive domain)
- Cognitive strategies (Cognitive domain)
- Motor skills (Psycho-Motor domain)
- Attitudes (Affective domain)

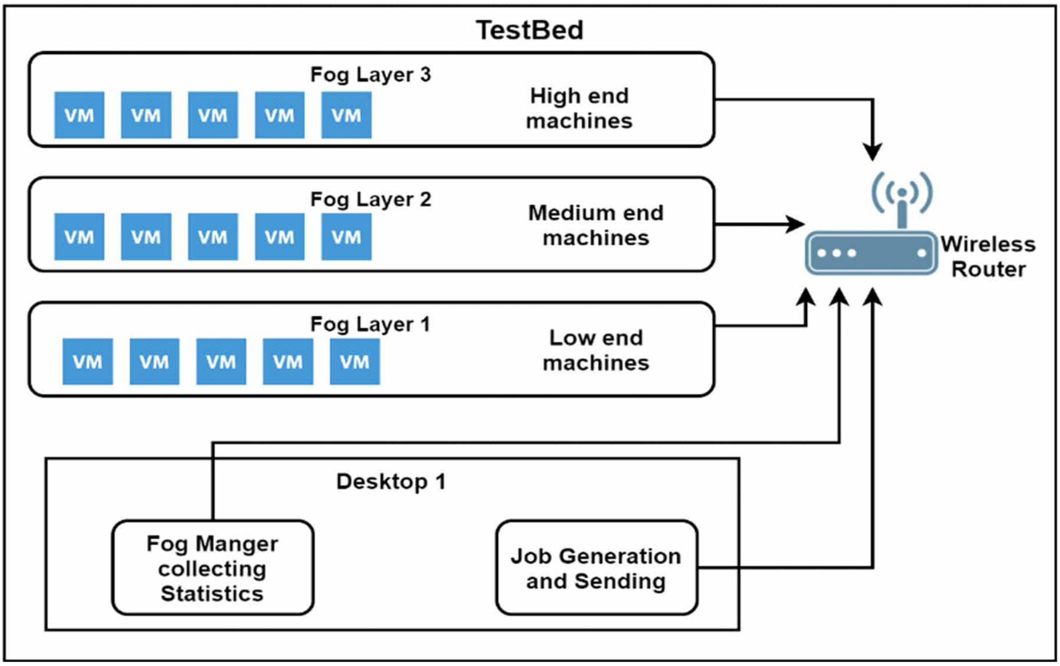
3.2.1. Gagné's Levels of Learning

According to Gagne, to achieve the five conditions of learning, students progress through nine levels of learning, and any teaching session should make sure the class plan is planned according to nine levels. The idea was that the nine levels of learning activate the five conditions of learning and thus, learning will be achieved.

Gagne's nine levels of learning are listed below:

1. Gain attention.
2. Inform students of the objective.
3. Stimulate recall of prior learning.
4. Present the content.
5. Provide learning guidance.
6. Elicit performance (practice).
7. Provide feedback.
8. Assess performance.
9. Enhance retention and transfer to the job.

Figure 3. Bloom's taxonomy



3.3 Maslow's Hierarchy of Needs

The basic idea for Maslow's hierarchy of needs is that students progress through a set of psychological needs to self-actualization (Figure 4).

Maslow's theory is all about building student/teacher relationships rather than lesson or curriculum structure. He states that if the teacher does not show passion, enthusiasm and empathy, it is challenging for the students to meet their needs despite having the best resources and lesson plans."

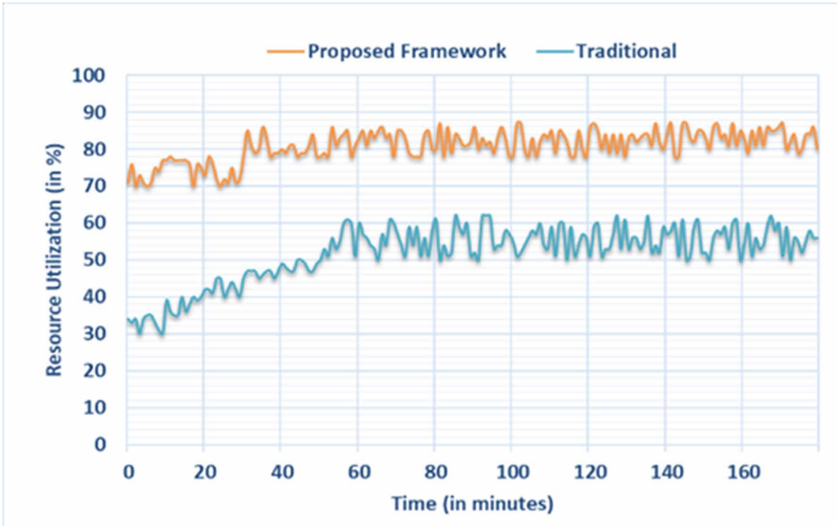
3.4 Howard Gardner's Multiple Intelligences

Howard Gardner is an American psychologist who was working as a professor at Harvard University. In 1983 he published "Frames of Mind" where he laid out his theory of "multiple intelligences".

The Intelligences proposed by Gardner are as follows:

1. **Linguistic Intelligence:** The ability to understand and express one's thoughts by writing and speaking in a language
2. **Mathematical Intelligence:** The intelligence to solve scientific, mathematical problems logically and to perform experiments.
3. **Musical Intelligence:** Skill to appreciate, compose and perform musical notes and the ability to understand the sound of the tone, pitch and rhythm.
4. **Bodily-Kinesthetic Intelligence:** Ability to solve problems by coordinating mind and body movements.
5. **Spatial Intelligence:** Being able to understand and use patterns in a wide or restricted area.
6. **Interpersonal Intelligence:** The capacity to perceive others desires, motivations and intentions.
7. **Intrapersonal Intelligence:** The potential of understanding self fears, feelings and motivations.

Figure 4. Maslow's hierarchy of needs



Gardner suggested that the intelligence compliment each other when students learn new skills and solve problems and rarely occur independently.

3.5 Kolb's Experiential Theory

In 1984, four-stage experiential learning theory was proposed by an American education theorist named David Kolb (Figure 5). It is built based on:

Learning is the process whereby knowledge is created through the transformation of experience.

Learning is fulfilled when all four stages have completed. Each step in the cycle both supported by and led into the next step. A learner may undergo the cycle any number of times so that he/she refine their understanding of the topic further.

4. RESEARCH PERSPECTIVES ON EDUCATION

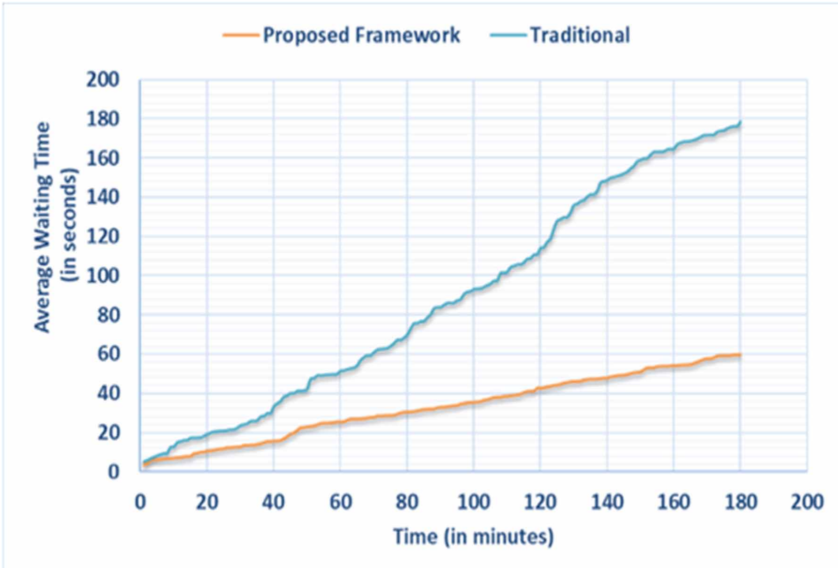
Educational research refers to a concerted effort to achieve a deeper understanding of the method of schooling, usually to increase its performance. This is an extension of the scientific method to research the problems of Education. The educational analysis is the application of the scientific method to the study of issues in Education. Therefore the steps of educational research are more or less comparable with those in the scientific method. The measures commonly found in education research are as follows: Research problem, Formulation of hypothesis, methods to be used, data collection, analysis and interpretation of data, and reporting the results.

This section review some educational research based on student academic performance, behaviour and emotions.

4.1 Students' Academic Performance

For an institution, academic success is critical for the positive results which will contribute to future job success. Arsenis and Flores (2019) study the relationship between the academic achievement of students and the possibility of gaining work experience as part of their undergraduate studies.

Figure 5. Kolb's experiential theory



Kostagiolas et al., (2019) provide evidence from the learning analytics to assess the effect of research satisfaction on academic self-efficacy and success of students. Academic achievement and student wellbeing are two ideal outcomes in phase-in learning and teaching. In a particular sense, the research in (Palos et al., 2019) measures the transient order of the link between educational success and two types of student progress. Path analysis models indicated that college grades could be considered a precedent of student participation and burnout, while wellbeing measures cannot be considered a precedent of academic success.

Budu et al., (2019) review the genesis of tension among midwife students in Ghana and its influence on their academic performance. The multivariate study showed that the responses of respondents when anxious had a significant effect on their academic success. However, having fewer holidays after accounting for the stressors had a substantial impact on the academic performance of the respondents. Meanwhile, the contact concept greatly enhances respondents who have had ample time to relax during their holidays. D'Alessio et al., (2019) examine the effect of critical thought on corporate Master of Business Administration (MBA) students ' academic achievement. Critical thinking has a positive impact on the average MBA student's academic success.

4.2 Students' Behavior

Student misbehaviour such as offensive speech, constant job evasion, clowning, dissatisfaction with instructional tasks, bullying peers, verbal abuse, instructor rudeness, defiance, and hatred, varying from rare to regular, moderate to severe, is a thorny problem in a daily classroom. Research has shown that school abuse has not only intensified over time but has also reduced academic performance and increased delinquency (Sun & Shek 2012).

Automatic review of the in-classroom behaviour of the students is useful for evaluating the teaching impact. Yang et al., (2020) aim to assess the degree of focus of the students concerning the instructor or quality of the instruction. Detect the faces of the students, map faces and evaluates the actions of the students, i.e. lift or bow faces and related instructor head orientations, teach material or not. In (Cantabella et al., 2019) case study undertaken at the Catholic University of Murcia, where student behaviour has been evaluated in the past four academic years based on learning modality, given the number of accesses to the Learning Management System, the resources used by students and their related behaviours.

The student behaviour analysis and prediction model focused on big data from the campus is developed in (Tu 2019), and the importance of big data produced by the behaviour of the campus students is analyzed. The behavioural data of the laws of consumption, living patterns and learning conditions of students are gathered, modelled, examined and excavated around the broad data sets, and the behaviour of the students is predicted and informed by the stratified model of behavioural characteristics of the students.

4.3 Students' Emotions

Through schooling and in all aspects of human life, feelings are of considerable value. It is understood that while people's backgrounds, the world in which they live, and the language they use vary, there are emotions that are regarded as universal. The students ' facial expressions were evaluated in (Tonguç & Ozkara 2020) and digitized in terms of feelings of dissatisfaction, disappointment, joy, anxiety, contempt, anger and surprise. The author also analyzed whether student emotions differed and how this difference was statistically meaningful based on their divisions, gender, lecture hours, machine position in the classroom, and style of lecture and session details.

Sahla and Kumar (2016) propose a deep learning method for emotion analysis. It focuses on students of a classroom and thus, understands their facial emotions. The neural networks of convolution estimate emotional groups with the highest likelihood as a consequence. An appraisal should include the user, for example, depending on the expected emotion; whether the students are satisfied and this class is interesting.

Concentration review of the students will help to strengthen the learning experience. Emotions are directly related and represent the attention of students closely. In (Sharma et al., 2019), a research program is introduced to assess the level of focus in real-time from the articulated facial emotions. The emotions conveyed was associated with the students ' concentration, and three distinct arousal levels (high, medium and low) were conceived. Bosch et al., (2016) using computer vision, learning analytics, and machine learning to predict the emotions of the students in a school computer lab's real-world environment. It succeeded in detecting fatigue, confusion, excitement, anger and focus in a way that was universal through pupils, time and demographics.

5. SENTIMENT ANALYSIS

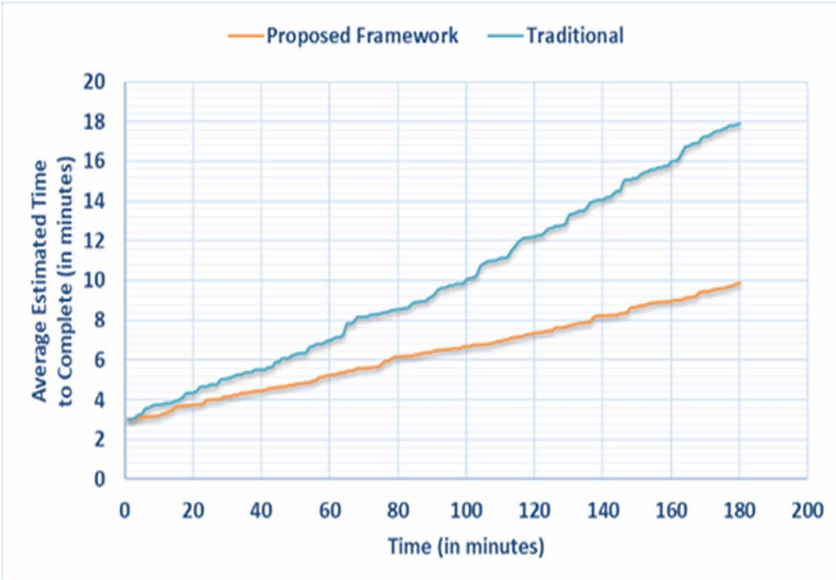
It is recognized as a sentiment analysis (SA) to understand and categorize the emotions of users from a section of the text into specific emotions. For example, emotions such as joyful, depressed, angry or optimistic, negative or neutral to decide the mood of the users concerning a particular subject or event. This method of research is also called opinion mining (with an emphasis on extraction) or affective rating.

Sentiments refer to attitudes, opinions, and emotions. In other words, the interpretations are emotional as opposed to objective truth. Different forms of sentiment analysis use various methods and approaches to recognize the sentiments found in a given text.

5.1 Approaches of Sentiment Analysis

SA has three significant levels of classification (Medhat et al., 2014). In essence, document-level, sentence-level, and aspect-level SA. Document-level SA is aimed at classifying an opinion document as reflecting a positive or negative viewpoint or sentiment. This finds the whole text to be a simple unit of knowledge (talking of one subject). Sentence-level SA is aimed at classifying the emotions conveyed in each sentence. The first step is to see if the statement is subjective or objective. If the sentence is subjective, it will be determined by Sentence-level SA if the sentence expresses positive or negative opinions. Aspect-level SA is directed at classifying sentiment for the individual characteristics

Figure 6. Types of sentiment analysis approaches



of individuals. The first step is to define the organizations and the facets thereof. The opinion holders will express different views on various aspects of the same organization.

Several algorithms, techniques are used to achieve sentiment analysis. Figure 6 shows the types of sentiment analysis approaches.

There are four necessary steps are used to analyze the sentiments D'Andrea et al., 2015). They are

1. **Data Collection:** The first phase of the sentiment analysis is to collect data from content created by users found in blogs, forums and social networks. These details are disorganized, articulated in multiple ways by the use of specific words, slangs, writing sense etc. Research by hand is almost impossible. Text mining and natural language modelling are also included in the retrieval and classification.
2. **Preprocessing:** Consists of cleaning before evaluating the extracted results. It detects and excludes non-textual contents and contents that are inappropriate for analysis.
3. **Sentiment Detection:** Reviews and opinions extracted sentences are examined. Sentences with political statements (opinions, views, and beliefs) are preserved, and sentences of impartial communication (facts, truthful information) discarded.
4. **Sentiment Classification**: In this stage, subjective sentences are categorized as positive, negative, good, bad; like hate, but they can be categorized by several points.

5.2 Applications of Sentiment Analysis

Sentiment Analysis can be used in many applications. Some of the applications are listed below:

- **Market and FOREX Rate Prediction:** Foreign Currency Exchange plays an essential role in the financial market for currency trading. The market is predicted by analyzing the twitter sentiments on the commercial market.
- **Box Office Prediction:** Success of the upcoming movie is predicted using sentiment analysis on youtube comments, tweets, blogs etc.

- **Business:** Demand and Distribution choices should be taken depending on the user's perceptions of the product. Via these observations, the company can check the quality of the service it offers. Growing decisions in the market can be taken based on timely sentiments available.
- **Market Intelligence:** Designed to satisfy four needs of business managers likes (a) Possibilities and threat Determination, (b) Identifying opponents, (c) Help to acquire competitors' progress, and (d) aid great marketing decision making.
- **Politics:** Debated political issues on online forums. The public figure's positive and negative influence can be identified by examining people's thoughts on social media.
- **Recommenders System:** When customers use it enthusiastically, the good or service itself would have pleasant emotions. If the user considers the scores or emotions, these items may be highly recommended to a potential customer. Analysis of opinion also plays a critical role in supporting the system.
- **Summarization:** This is time-consuming for a reader to read all of the opinions about a particular institution and then judge. SA should give us every organization's general idea for a given period.
- **Government Intelligence:** The growth in violent behaviour can be monitored for tracking the sources. For making policies, the sentiments of people can be studied. This can be used for evaluating people's attitude about every SA conflict.
- **Education:** Analysis student behaviour, emotions based on the student feedback. In this educational framework, students provide feedback on Twitter at any time or at different time slots, as determined by the professor, to ensure that students observe the tempo of the lecture and provide assistance when faced with difficulties.

5.3 Research on Sentiment Analysis in Education

This section explains how sentiment analysis is used in Education. Rajput et al., (2016) suggested the sentiment analysis offered by students at the end of a course on faculty assessment. A Knime workflow was developed using its text processing feature for feel analysis of input from students. This method suggests a sentiment score measurement to identify the feedback as either positive, negative or neutral.

Nasim et al., (2017) propose a combination of machine learning and lexicon-based approaches to student input emotion analysis. The textual input, usually obtained at the end of a course, offers valuable insights into the general level of teaching and proposes useful ways to enhance teaching methods.

Aung and Myo (2017) plan to systematically evaluate the text feedback of the students using a lexicon-based approach to forecast the teaching performance levels. A dictionary of English sentiment terms is generated to get the polarity of terms as a lexical source.

Sujata Rani and Parteek Kumar (2017), proposed sentiment analysis system using performing temporal sentiment and emotion analysis of multilingual to enhance teaching and learning based on student feedback on teacher performance and course satisfaction.

Sultana et al., (2018) proposes a model based on Deep Learning approach to perform sentiment analysis on Educational data.

Featherstone and Botha (2015) reports on the teachers' experiences those who are participating in a professional development programme through sentiment analysis. It was inferred that those teachers were happy with the training, felt the course relevance strongly.

Sivakumar and Reddy (2017) using R tools, from Twitter API student feedbacks were collected for analysis. K-mean clustering and Naïve Bayes classification algorithm was used to perform sentiment analysis. The outcome of the proposed work will help the students to improve their learning skills and helps the academicians to enhance their teaching methodologies so that educational institutions can resolve the students' problems and attain their motto.

Nasim et al. (2017) proposes a combination of machine learning and lexicon-based methods to analyze the students feedback using sentiment analysis. With experimental results, it was inferred the performance of the proposed model is better than other methods.

Barron-Estrada et al. (2019) This paper presents a sentiment analyzer to detect student sentiment and/or emotions by recognizing the polarities and emotions on learning using textual phrases that are written in Spanish. The results obtained is 88.26% accuracy.

6. CONCLUSION

Education is concerned with methods of learning and teaching that facilitates the process of learning, or the acquisition of knowledge, skills, values, beliefs, and habits in the learning environments setup. The theories on education and learning process state that the motive of education is not that an individual grows successful as a professional but also as a complete person. That is, the student should grow himself in knowledge, behaviour, attitude, skills, emotions etc. through the learning process. The theories also state that education alone does not improve students skill, knowledge, behaviour, feelings, etc.,but also the environment and experiences he undergoes. This survey presents a brief background of Education with different learning theories. This paper describes various researches on Education using multiple methods of education data mining and sentiment analysis. It mainly focuses on student academic performance, behaviour and emotions in the academic environment and processes that assess these factors. It is inferred from the survey that education grooms a person in knowledge,skillset and emotions. The advances of computer technologies and research domains like educational data mining and sentiment analysis help out in identifying the issues in achieving the motive of education and satisfying the theories of learning by supporting in perfect decision making using machine learning algorithms.

REFERENCES

Algarni, A. (2016). Data Mining in Education. *International Journal of Advanced Computer Science and Applications*, *7*(6).

Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014). Sentiment Analysis: Towards a Tool for Analysing Real-Time Students Feedback. *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, 419-423.

Altrabsheh, N., Gaber, M. M., & Cocea, M. (2013). SA-E; Sentiment analysis for education. *Frontiers in Artificial Intelligence and Applications*, *255*, 353–362.

Arsenis, P., & Flores, M. (2019). *Student academic performance and professional training year* (Vol. 30). International Review of Economics Education.

Aung, K. Z., & Myo, N. N. (2017). Sentiment analysis of students' comment using the lexicon-based approach. *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS),* 149-154. doi:10.1109/ICIS.2017.7959985

Barron-Estrada, M. L., Zatarain-Cabada, R., & Oramas-Bustillos, R. (2019). Emotion Recognition for Education using Sentiment Analysis. *Research in Computing Science*, *148*(5), 71–80. doi:10.13053/rcs-148-5-8

Bloom. (n.d.). *Taxonomy of educational objectives*. New York, NY: David Mckay.

Bosch, N., D'Mello, S. K., Baker, R. S., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., & Zhao, W. (2016). Detecting StudentEmotions in Computer-Enabled Classrooms. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 4125–4129.

Budu, H. I., Abalo, E. M., Bam, V., Budu, F. A., & Peprah, P. (2019). A survey of the genesis of stress and its effect on the academic performance of midwifery students in a college in Ghana. *Midwifery*, *73*, 69–77. doi:10.1016/j.midw.2019.02.013 PMID:30903921

Cantabella, M., Martínez-España, R., Ayuso, B., Yáñez, J., & Muñoz, A. (2019). Analysis of student behavior in learning management systems through a Big Data framework. *Future Generation Computer Systems*, *90*, 262–272. doi:10.1016/j.future.2018.08.003

D'Alessio, F. A., Avolio, B. E., & Charles, V. (2019). Studying the impact of critical thinking on the academic performance of executive MBA students. *Thinking Skills and Creativity*, *31*, 275–283. doi:10.1016/j.tsc.2019.02.002

D'Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches Tools and Applications for Sentiment Analysis Implementation. *International Journal of Computers and Applications*, *125*(3).

Featherstone, C., & Botha, A. (2015, May). Sentiment analysis of the ICT4Rural Education teacher professional development course. In *2015 IST-Africa Conference* (pp. 1-12). IEEE.

Kostagiolas, P., Lavranos, C., & Korfiatis, N. (2019). *Learning analytics: Survey data for measuring the impact of study satisfaction on students' academic self-efficacy and performance* (Vol. 25). Data in Brief.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. doi:10.1016/j.asej.2014.04.011

Mining, T. E. D. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *Proceedings of the conference on advanced technology for Education*.

Nasim, Z., Rajput, Q., & Haider, S. (2017, July). Sentiment analysis of student feedback using machine learning and lexicon-based approaches. In *2017 international conference on research and innovation in information systems (ICRIIS)* (pp. 1-6). IEEE.

Nasim, Z., Rajput, Q., & Haider, S. (2017). Sentiment analysis of student feedback using machine learning and lexicon-based approaches. *2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*, 1-6. doi:10.1109/ICRIIS.2017.8002475

Paloș, R., Maricuţoiu, L. P., & Costea, I. (2019). engagement and student burnout: A cross-lagged analysis of a two-wave study. *Studies in Educational Evaluation*, *60*, 199–204. doi:10.1016/j.stueduc.2019.01.005

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). *The Control-Value Theory of Achievement Emotions: An Integrative Approach to Emotions in Education*. Emotion in Education.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist*, *37*(2), 91–105.

Rajput, Q., Haider, S., & Ghani, S. (2016). Lexicon-based sentiment analysis of teachers evaluation. *Applied Computational Intelligence and Soft Computing*, *2016*, 1–12. doi:10.1155/2016/2385429

Rani, S., & Kumar, P. (2017). A sentiment analysis system to improve teaching and learning. *Computer*, *50*(5), 36–43. doi:10.1109/MC.2017.133

Sahla, K. S., & Kumar, T. S. (2016). Classroom Teaching Assessment Based on Student Emotions. *The International Symposium on Intelligent Systems Technologies and Applications*, 475-486. doi:10.1007/978-3-319-47952-1_37

Sharma, P., Esengönül, M., Khanal, S. R., Khanal, T. T., Filipe, V., & Reis, M. J. C. S. (2019). Student Concentration Evaluation Index in an E-learning Context Using Facial Emotion Analysis. *International Conference on Technology and Innovation in Learning, Teaching and Education*, 529-538. doi:10.1007/978-3-030-20954-4_40

Sivakumar, M., & Reddy, U. S. (2017, November). Aspect based sentiment analysis of students opinion using machine learning techniques. In *2017 International Conference on Inventive Computing and Informatics (ICICI)* (pp. 726-731). IEEE. doi:10.1109/ICICI.2017.8365231

Sultana, J., Sultana, N., Yadav, K., & AlFayez, F. (2018, April). Prediction of sentiment analysis on educational data based on deep learning approach. In *2018 21st Saudi Computer Society National Computer Conference (NCC)* (pp. 1-5). IEEE. doi:10.1109/NCG.2018.8593108

Sun, R. C. F., & Shek, D. T. L. (2012). *Student Classroom Misbehavior: An Exploratory Study Based on Teachers' Perceptions*. Developmental Issues in Chinese Adolescents.

Tonguç, G., & Ozkara, B. O. (2020). *Automatic recognition of student emotions from facial expressions during a lecture* (Vol. 148). Computers & Education.

Tu, L. (2019). Analysis and Prediction Method of Student Behavior Mining Based on Campus Big Data. *International Conference on Advanced Hybrid Information Processing*, 363-371. doi:10.1007/978-3-030-36405-2_36

Yang, B., Yao, Z., Lu, H., Zhou, Y., & Xu, J. (2020). In-classroom learning analytics based on student behavior, topic and teaching characteristic mining. *Pattern Recognition Letters*, *129*, 224–231. doi:10.1016/j.patrec.2019.11.023

M. Amala Jayanthi is working as assistant professor in the Department of Computer Applications, Kumaraguru College of Technology, Coimbatore. She has 7 years of experience as academician. She received her Master of philosophy in Department of Computer Science, St. Joseph's College (Autonomous), Bharathiyar University, Tiruchirappalli, Tamil Nadu, India. She received her MCA degree from Kalasalingam University. Her current area of research is Data mining. She has published 11 research papers in various reputed journals.

I. Elizabeth Shanthi (PhD) is working as professor in the Department of Computer Science Avinashilingam Institution for Home Science and Higher Education for women, Avinashilingam University, Coimbatore, Tamil Nadu, India. She has 27 years of experience as academician. Her area of expertise is Data mining and Information Retrival. She has 21 published research papers in various reputed journals.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 4733550, 12 pages
https://doi.org/10.1155/2022/4733550

Research Article

Quest_SA: Preprocessing Method for Closed-Ended Questionnaires Using Sentiment Analysis through Polarity

Amala Jayanthi M [1] **and Elizabeth Shanthi I** [2]

[1]*Department of Computer Applications, Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India*
[2]*Department of Computer Science, Avinashilingam Institution for Home Science and Higher Education for Women, Avinashilingam University, Coimbatore, Tamil Nadu, India*

Correspondence should be addressed to Amala Jayanthi M; amalajayanthi.m.mca@kct.ac.in

Received 29 April 2022; Revised 15 June 2022; Accepted 18 July 2022; Published 20 September 2022

Academic Editor: M. Praveen Kumar Reddy

Sentiment analysis is a prominent research topic in natural language processing, with applications in politics, news, education, product review, and other sectors. Especially in the education sector, sentiment analysis can assist educators in finding students' feelings about a course on time, altering the teaching plan appropriately and timely to improve the quality of education and teaching. For students, the sentiment analysis can identify emotions, academic performance, behaviour, and so on; the primary purpose of this research paper is to analyze students' emotions, self-esteem, and efficacy based on closed-ended questionnaires. This paper proposes Quest_SA, which uses the sentiment analysis technique to identify students' emotions based on the answer provided by a closed-ended questionnaire. The polarity value is assigned for each questionnaire scale. The students' responses are then gathered using a closed-ended questionnaire, and the student's emotions are classified using a polarity-based method of sentiment analysis. Finally, sentiment scores and emotion variance were used to evaluate the outcomes. According to the sentiment ratings, students have favourable sentiments and emotions such as unhappy, somewhat happy, and happy. The real-world closed-ended questionnaires such as emotional intelligence, Eysenck, personality, self-determination scale, self-efficacy, Rosenberg's self-esteem, positive and negative affect schedule, and Oxford happiness questionnaires were used to examine the academic performance with the proposed sentiment analysis. This study inferred that the proposed sentiment analysis pre-processing method with polarity scores is as accurate as the standard value calculation.

1. Introduction

Sentiment analysis is a technique for detecting polarity and recognizing emotion toward a certain object, such as a person, a concept, or an activity. The purpose of sentiment analysis is to determine people's opinions, identify the emotions they express, and categorize them as positive, negative, or neutral. Natural language processing (NLP) and machine learning (ML) techniques are used by sentiment analysis systems to identify, retrieve, and synthesize information and opinions from large amounts of text [1].

In general, sentiment analysis was done at three levels: document, sentence, and aspect. Document Level Sentiment Analysis discovers the user sentiments by evaluating the entire document. The goal of sentence-level research is to establish the polarity of individual sentences rather than the entire document; as a result, it is more precise. Finally, aspect-level sentiment analysis identifies elements or attributes mentioned in reviews and categorizes users' reactions to them. The architecture of a broad sentiment analysis system is shown in Figure 1.

The whole system may employ a set of lexicons and linguistic resources. The document analysis module is a critical component of the system design since it employs linguistic resources to annotate preprocessed documents with sentiment annotations. The system's output—positive, negative, or neutral—is represented by annotations in several visualization tools. Depending on the sentiment

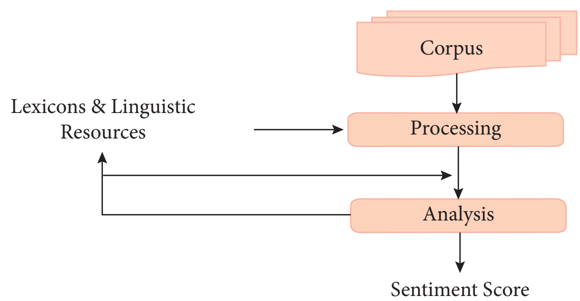


Figure 1: General sentiment analysis architecture[2].

analysis form, annotations may be utilized in various ways. For example, in document-based sentiment analysis, annotations may be applied to the entire document; in sentence-based sentiment analysis, annotations can be applied to specific sentences; and in aspect-based sentiment analysis, annotations can be applied to certain subjects or entities.

Sentiment analysis has been used in various settings to achieve a variety of goals, most notably in professional and economic networks. A few examples of well-known sentiment analysis business applications include product and service reviews [3], financial marketing approaches [4], and customer relationship management [5]. The most common use of sentiment analysis in social media apps is to analyze a company's reputation on Twitter or Facebook [6] and investigate people's reactions to a crisis, for example, COVID-19 [7]. Another important application area is politics [8], where sentiment research might aid candidates in their election campaigns.

Sentiment analysis and opinion mining have got a lot of attention in the educational community [9]. Unlike the previously stated sectors of social and commercial networks, which focus on a single user, education sentiment analysis research covers a variety of views, including teachers/instructors, students/learners, decision-makers, and institutions. Sentiment analysis is largely used to improve teaching, management, and assessment by examining learners' attitudes and behaviour toward courses, platforms, institutions, and teachers.

Sentiment analysis is utilized to investigate the relationship between learners' sentiments and drop-out rates in massive open online courses and the relationship between performance and retention and learners' emotions [10]. Finally, sentiment analysis has examined several teacher-related aspects expressed in student reviews or comments on discussion forums in terms of teacher viewpoints [11].

Students are frequently obliged to engage in postcourse questionnaires at the end of each academic term to obtain information about their experiences. This procedure allows teachers and administrators to review student assessments and improve learning processes. There are both closed- and open-ended questions on the survey. Closed-ended questions, frequently used in Likert-scale inquiries, try to capture students' evaluations in numerical ratings. Students can provide written comments or ideas in response to open-ended questions, which reflect their personal views and perceptions. This paper considers the closed-ended questions for identifying students' emotions using sentiment analysis. The students' responses are collected using a closed-ended questionnaire, and the students' emotions are specified using a polarity-based sentiment analysis algorithm. The outcomes were assessed using sentiment scores and emotion variance. According to the sentiment ratings, students have positive sentiments and emotions such as unhappy, somewhat happy, and happy.

The remainder of this study paper is structured as follows: the research backdrop is described in Section 2, which includes sentiment analysis and a questionnaire. After that, in Section 3, the recommended methodology is explained. Finally, in Section 4, the proposed work's performance is evaluated using standard questionnaires such as the Oxford happiness inventory, self-determination scale, Rosenberg's self-esteem, self-efficacy, emotional intelligence, Eysenck personality questionnaire, and positive and negative affect schedule, and the conclusion and future work of this research work are presented.

2. Related Work

2.1. Sentiment Analysis. Sentiment analysis [12] evaluates emotional representation through language that comprises acquiring dynamic data, processing and analyzing data, and classifying a piece of text. The three main sentiment analysis tasks are facial expression identification, polarity detection, and affective computing [13]. Text sentiment analysis is a realistic approach for emotion mining in natural language processing widely used in public opinion monitoring, artificial intelligence, and corporate analytics. The three primary methods for text sentiment analysis are a machine learning-based technique, a dictionary-based approach, and a hybrid approach [14].

In machine learning-based techniques, sentiment classifiers are trained using a prelabeled data set. A classifier can be created to determine the polarity of textual inputs using methods such as naive Bayes, support vector machine, maximum entropy, and Word2vec, which are commonly used in sentiment analysis [15]. The dictionary-based approach uses a predeveloped lexicon, which includes the contradiction of words or phrases, to compute sentiment ratings and detect the polarity of a given text. The sentiment score is based on open source or bespoke sentiment dictionaries and can be computed using numerous semantic criteria [16]. A hybrid approach to sentiment classification combines machine learning and dictionary approaches. In general, the machine learning-based method is more effective although it takes a long time to classify the data [17]. The dictionary-based technique, on the other hand, has the advantage of not requiring any training data to determine sentiment and is substantially faster than machine learning in terms of computing time.

Customer product review [18], sale predictions [19], social media data [20], sarcasm detection [21], and the economic domain [22] are just a few examples of where sentiment analysis has been used. In the subject of education, sentiment analysis has recently gained interest. Reference

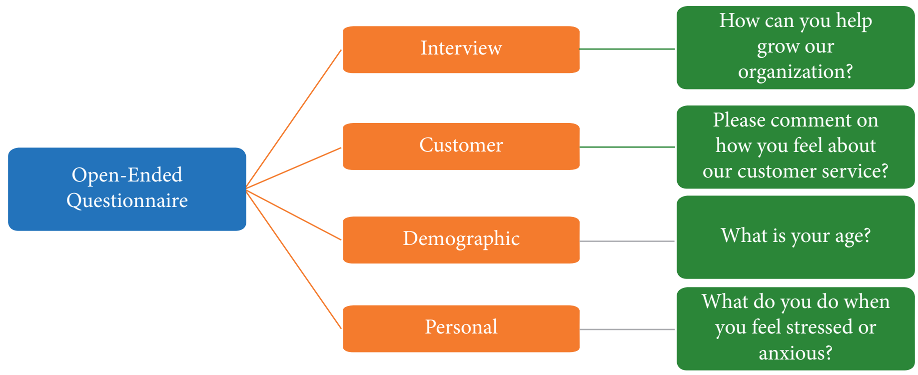


FIGURE 2: Example of the open-ended questionnaire.

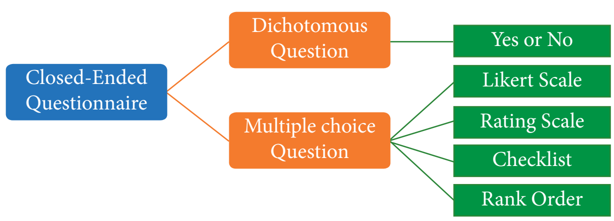


FIGURE 3: Types of closed-ended questions.

[23] employed a lexicon-based approach to judging document-level polarity on students' feedback to evaluate teachers.

Reference [24] introduced sentiment analysis provided by students at the end of a teacher evaluation course. The text processing capability of KNIME was utilized to build a pipeline for analyzing student feelings. This method recommends categorizing feedback as good, negative, or neutral using a sentiment score. Reference [25] proposed a hybrid technique for analyzing student input emotions that blends machine learning and lexicon-based methodologies. Textual feedback, usually given at the end of a course, provides useful insights into the general level of teaching and suggests practical ways to enhance teaching methods. Reference [26] planned to evaluate students' text feedback and estimate instructional success levels using a lexicon-based technique. A lexicon of English sentiment phrases is built to get the polarity of terms as a linguistic source. In a sentiment-based eSystem,(i) for film reviewing, client happiness is measured using sentiment analysis with hybrid fuzzy and deep neural network [27], (ii) for modern business, knowledge discovery and sentiment analysis is used [28]. Selection for the best SVM hyperparameter values is done by applying natural optimizing techniques [29]. (iii) for non-traditional learning, expansion of hybrid reality-based education is done [30].

2.2. Open-Ended Questionnaire. Open-ended questions are survey questions that allow respondents to respond in an open text format, conveying their complete understanding, feelings, and knowledge. It implies that the answer to this question is not limited to a few options. Open-ended questions are commonly used in qualitative market research. A question with an open-ended response allows the viewer to answer depending on their knowledge and experience. The viewer's detailed and elaborate knowledge leaves the potential for additional discussion and improvement. An open-ended question provides opportunities for both the researcher and the respondent to learn.

Figure 2 shows some examples of open-ended questionnaires.

The open-ended questionnaire has many merits, but it is difficult to analyze and organize the data into reports. Too many questions can directly harm the response rate. Moreover, the open-ended questionnaire may provide irrelevant information.

2.3. Closed-Ended Questionnaire. A questionnaire is a research tool that consists of questions or other prompts designed to gather data from a respondent. There are two types of questionnaires: structured and unstructured questionnaires. Quantitative data were collected via structured questionnaires. Quantitative questionnaires are used to evaluate or verify the accuracy that has already been developed. The questionnaire is meticulously constructed and designed to collect precise data. It also starts a formal investigation, contributes data, double-checks previously gathered data, and aids in invalidating any previous idea. Unstructured surveys are used to gather qualitative information. For example, qualitative questionnaires are used when collecting exploratory data to prove or reject a theory. They employ a minimal structure and a few branching questions, but nothing restricts a respondent's options. To acquire specific responses from people, the questions are more open-ended. This research work considers structured quantitative questionnaires (closed-ended questionnaires). An investigation of the association between self-esteem and students' academic performance was done in [31]. The authors of [32] worked on research contemplated on educational data mining.

Closed-ended questions, such as "yes" or "no" or multiple-choice questions, require respondents to choose from a limited set of predefined responses. Closed-ended inquiries are frequently used to gather statistical data from

Table 1: Sample questions.

Questionnaire	Sample questions	Scale
Emotional intelligence	1. I realize immediately when I lose my temper	Not at all
	2. I can reframe bad situation quickly	Rarely
	3. I am always able to motivate myself to do difficult tasks	Sometimes
	4. I am always able to see a thing from the other person's viewpoint	Often
	5. I am an excellent listener	Very often
Eysenck personality	1. Does your mood often go up and down?	
	2. Do you take much notice of what people think?	Yes
	3. Are you a talkative person?	No
	4. If you say you will do something, do you always keep your promise, no matter how inconvenient it might be?	
	5. Do you ever feel just miserable for no reason?	
Self-determination scale	1A. I always feel like I choose the things I do	1
	1B. I sometimes feel that it is not really me choosing the things I do	2
	2A. My emotions sometimes seem alien to me	3
	2B. My emotions always seem to belong to me	4
	3A. I choose to do what I have to do	5
	3B. I do what I need to do, but I do not feel like it is really my choice	
General self-efficacy	1. I can always manage to solve difficult problem if I try hard enough.	Very slightly or not at all
	2. If someone opposes me, I can find the means and ways to get what I want	A little
	3. It is easy for me to stick to my aim and accomplish my goals	Moderately
	4. I am confident that I could deal efficiently with unexpected events	Quite a bit
	5. Thanks to my resourcefulness, I know how to handle unforeseen situations	Extremely
Rosenberg's self-esteem	1. On the whole, I am satisfied with myself	Strongly agree
	2. At times, I think I am no good at all	Agree
	3. I feel that I have a number of good qualities	Disagree
	4. I am able to do things as well as most other people.	Strongly disagree
	5. I feel I do not have much to be proud of	
Positive and negative affect schedule	1. Interested	Very slightly or not at all
	2. Distressed	A little
	3. Excited	Moderately
	4. Upset	Quite a bit
	5. Strong	Extremely
Oxford happiness	1. I do not feel particularly pleased with the way I am (R)	Strongly disagree
	2. I am intensely interested in other people	Moderately disagree
	3. I feel that life is very rewarding	Slightly disagree
	4. I have very warm feelings towards almost everyone	Slightly agree
	5. I rarely wake up feeling rested (R)	Moderately agree
		Strongly agree

responders. It can take various shapes, but they are all driven by the requirement for respondents to have particular choices. Figure 3 depicts many sorts of closed-ended questions.

Table 1 shows the sample questions and the Likert scale used in each questionnaire.

The Oxford happiness questionnaire was developed by psychologists [33] at Oxford University. In the Oxford questionnaire, the (R) indicates reverse scoring. For example, if the student gives "1," cross it out and change it to "6." The emotional intelligence questionnaire is a self-evaluation tool. Self-awareness, self-regulation, motivation, empathy, and social skills are the five characteristics that characterize emotional intelligence, according to [34]. The self-esteem of an individual is assessed using Rosenberg's self-esteem scale. The score of negative question items is

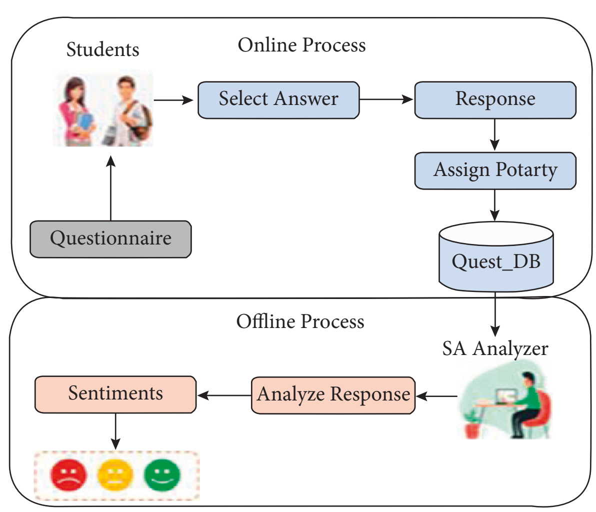


Figure 4: Quest_SA architecture.

Table 2: Polarity score.

Questionnaire	Seale	Standard value	Proposed polarity value
Emotional intelligence	Not at all	1	−2
	Rarely	2	−1
	Sometimes	3	0
	Often	4	1
	Very often	5	2
Eysenck personality	Yes	1	1
	No	0	−1
Self-determination scale	1	1	−2
	2	2	−1
	3	3	0
	4	4	1
	5	5	2
General self-efficacy	Very slightly or not at all	1	−2
	A little	2	−1
	Moderately	3	0
	Quite a bit	4	1
	Extremely	5	2
Rosenberg's self-esteem	Strongly agree	4	2
	Agree	3	1
	Disagree	2	−1
	Strongly disagree	1	−2
Positive and negative affect schedule	Very slightly or not at all	1	−2
	A little	2	−1
	Moderately	3	0
	Quite a bit	4	1
	Extremely	5	2
Oxford happiness	Strongly disagree	1	−3
	Moderately disagree	2	−2
	Slightly disagree	3	−1
	Slightly agree	4	1
	Moderately agree	5	2
	Strongly agree	6	3

inverted for analysis such that the positive and negative things have the same meaning. The final test result might be between 10 and 40. A person with a score of less than 14 has a problem with low self-esteem and needs assistance. The Eysenck personality test is a self-reporting tool [35]. It has 48 items: 12 for each of the personality traits of neuroticism, extraversion, and psychoticism and 12 for the lying scale. "Yes" or "no" is the binary response to each inquiry. Each dichotomous item was given a value of 1 or 0, with a maximum score of 12 and a minimum of 0. The self-determination scale (SDS) was developed to examine how self-determined people perform individually. It is thus regarded as a reasonably stable feature of people's personalities that reflects (1) increased awareness of their feelings and sense of self and (2) a sense of control over their behaviour. The general self-efficacy scale is a 10-item psychometric scale that assesses optimistic self-beliefs in one's ability to cope with various life challenges. Positive and negative affect schedule (PANAS) is a scale of several words that express feelings and emotions. The overall score is computed by adding 10 positive items together and then 10 negative items. For both sets of objects, the scores range from 10 to 50. A greater total positive score suggests a stronger beneficial influence. A lower total negative score suggests a lesser level of negative impact.

3. Methodology

Sentiment analysis is a computational study that evaluates individuals' thoughts, assessments, and opinions regarding persons, situations, entities, concepts, activities, and items and their characteristics. Its goal is to find underlying opinions on a specific entity automatically. Sentiment analysis is mainly used for commercial applications such as product reviews, recommendations, marketing analysis, and public relations [36]. In the field of education, sentiment analysis is the process of determining a student's feelings. In education, sentiment analysis can help with learning process improvement, performance improvement, study discontinuance reduction, teaching process improvement, and course satisfaction.

Emotion is commonly defined as a person's mental state, including attitudes, feelings, and actions. Nowadays, public sentiment on a particular context can be easily known by extracting the opinions from a wealth of

Table 3: Questionnaire details.

Questionnaire	No. of questions	Scale	Score calculations	Result	Result with polarity
Emotional intelligence connects a person's knowledge process to their emotional processes	15	1 = Not at all 2 = Rarely 3 = Sometimes 4 = Often 5 = Very often	Sum all scale values for each item	<34 = Low 35–55 = Average >56 = High	−ve Score = low 0 = Average +ve Score = high
Eysenck personality measures the personality domain	48	1 = Yes 0 = No	Sum all scale values of psychoticism (PM), extroversion (En), and neuroticism (Nm)	The biggest value of psychoticism, extroversion, and neuroticism (Pm, En, and Nm)	The positive value of psychoticism, extroversion, and neuroticism (Pm, En, and Nm)
Self-determination scale assesses individual differences in the extent to which people tend to function in a self-determined way	10	1 2 3 4 5	Sum all result scale values and divide by 10	<3 = Low >3 = High	+ve = high −ve = Low
General self-efficacy is a self-report measure of self-efficiency	10	1 = Very slightly or not at all 2 = A little 3 = Moderately 4 = Quite a bit 5 = Extremely	Sum all result scale values	<20 = Low >20 = High	−ve Score = low +ve Score = high
Rosenberg's self-esteem measures global self-worth	10	4 = Strongly agree 3 = Agree 2 = Disagree 1 = Strongly disagree	Sum all result scale values	<14 = Low 15–25 = Normal >26 = High	−ve Score = low 0 = Normal +ve Score = high
Positive and negative affect schedule is a self-reported measure of affect	20	1 = Very slightly or not at all 2 = A little 3 = Moderately 4 = Quite a bit 5 = Extremely	Sum all positive (PS) and negative (NS) affect scale values	PS > NS = positive NS > PS = negative PS = NS = false	PS = +ve score positive NS = +ve score negative
Oxford happiness is used to predict the happiness score of the person	29	1 = Strongly disagree 2 = Moderately disagree 3 = Slightly disagree 4 = Slightly agree 5 = Moderately agree 6 = Strongly agree	Sum all scale values and divide by the total number of questions	1–2: Not happy 2–3: Somewhat unhappy 3–4: Not particularly happy or unhappy 4: Somewhat happy 4–5: Happy 5–6: Very happy 6: Too happy	−ve Score = happy 0 = Moderately happy +ve Score = unhappy

Table 4: Emotional intelligence details.

Emotional intelligence			
Standard evaluation		Proposed evaluation	
Result	Count	Result	Count
Low	10	Low Average High	10 0 0
Average	290	Low Average High	10 265 15
High	700	Low Average High	0 0 700

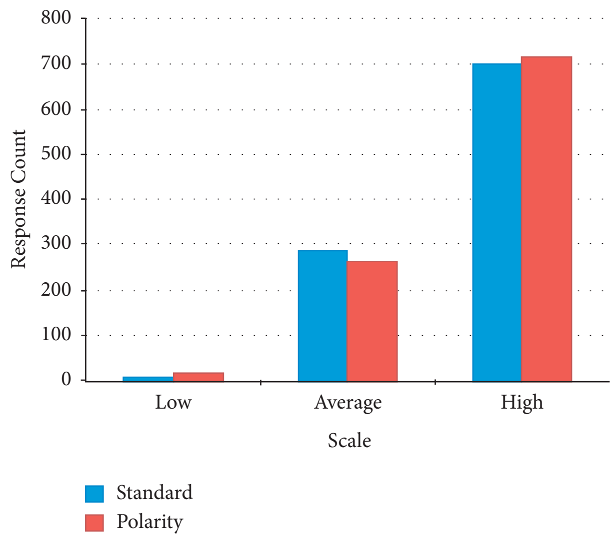


Figure 5: Emotional intelligence (standard vs. proposed).

publicly available information on platforms such as Facebook, Youtube, Twitter, and Instagram.

Emotion detection, Reddit, Twitter, and others are becoming increasingly popular as a new study horizon in NLP. It could also be used in health services (as a tool for psychoanalysis), education (identifying learner dissatisfaction), and other fields [37]. This paper proposes Quest_SA to find students' affective traits using polarity-enabled sentimental analysis in a closed-ended questionnaire. Figure 4 shows the architecture of the proposed Quest_SA.

The proposed Quest_SA contains two phases: online and offline processes. The students are requested to take an online closed-ended questionnaire in the online phase.

The following seven kinds of questionnaires are used: emotional intelligence, Eysenck personality, self-determination scale, self-efficacy, Rosenberg's self-esteem, positive and negative affect schedule, and Oxford happiness. The selected scale value is converted into a polarity value for each question. Table 2 shows the polarity assignment for the questionnaire scale.

In the offline phase, the sentiment analyzer uses the polarity value to predict the students' emotions. The polarity values are given based on the positive and negative sense. The positive Likert scale is given a positive score, and negative Likert scale is assigned a negative score.

4. Proposed QUEST_SA: Questionnaire Evaluation Using Sentiment Analysis

In this section, the performance of the proposed work is analyzed. Seven real-world questionnaires are used in this experiment. Table 3 shows the details of the questionnaire and the calculation in determining the result is shown.

The same standard calculation given in Table 3 is calculated for each questionnaire with the respective polarity value shown in Table 2. The result is given based on the polarity: if the result has negative polarity, then the scale is low; else, if the result has positive polarity, then the scale is high; and if the result is zero, then the scale is moderate. For instance, in Rosenberg's self-esteem, if the score is negative, the result is low self-esteem, and if the score is positive, the result is high self-esteem.

5. Results and Discussions

This section evaluates the performance of the proposed through experiments. The research work uses seven kinds of questionnaires such as emotional intelligence (EI), Eysenck personality (EP), self-determination scale (SDS), general self-efficacy (GSE), Rosenberg's self-esteem (RSE), positive and negative affect schedule (PNAS), and Oxford happiness (OH) and collects response from 1,000 students. The collected response was analyzed based on the standard and polarity-based evaluation. Finally, the obtained results are calculated and evaluated using MAE (mean absolute error) and accuracy.

Table 4 shows the emotional intelligence standard and the proposed polarity-based results. In addition, it shows the comparison of the result for all possible results. The result shows that the percentage of result deviation is very low between the standard evaluation and the proposed evaluation. Figure 5 shows the EI questionnaire scale value: low, average, and high. Table 5 gives the sample code of SentimentAnalyzer.

Table 6 shows the MAE and accuracy comparison of EI for different responses. Again, the lower number of the responses (200 and 400) produces a lower error.

Table 7 shows the Eysenck personality standard and the proposed polarity-based results. Figure 6 shows the questionnaire scale value for psychoticism, extroversion, and neuroticism. The standard and polarity evaluation produce the same result for all scale values.

Table 8 shows the MAE and EP accuracy comparison for a different number of responses. Again, the result produces zero error and 100% accuracy for all different numbers of responses.

Table 9 shows the self-determination scale standard and the proposed polarity-based results. Figure 7 shows the SDS questionnaire scale value: low and high.

Table 10 shows the MAE and SDS's accuracy comparison for a different number of responses.

Table 11 shows the MAE and accuracy comparison of GSE for a different number of responses.

Table 12 shows the general self-efficacy standard and proposed polarity-based result. Again, the standard and polarity evaluation results for low-scale values. Figure 8 shows the GSE questionnaire scale value low and high.

Table 13 shows Rosenberg's self-esteem standard and proposed polarity-based result. Figure 9 shows the RSE questionnaire scale value: low, normal, and high.

Table 14 shows the MAE and accuracy comparison of RSE for a different number of responses.

Table 15 shows the positive and negative affect schedule standard and the proposed polarity-based result. The

Table 5: Sample code of SentimentAnalyzer.

```
/*
* To change this license header, choose License Headers in Project Properties.
* To change this template file, choose Tools | Templates
* and open the template in the editor.
*/
package servlet1;
import java.util.Properties;
import org.ejml.simple.SimpleMatrix;
import edu.stanford.nlp.ling.CoreAnnotations;
import edu.stanford.nlp.neural.rnn.RNNCoreAnnotations;
import edu.stanford.nlp.pipeline.Annotation;
import edu.stanford.nlp.pipeline.StanfordCoreNLP;
import edu.stanford.nlp.sentiment.SentimentCoreAnnotations;
import edu.stanford.nlp.trees.Tree;
import edu.stanford.nlp.util.CoreMap;
/**
*
* @author jayanthi
*/
public class SentimentAnalyzer
{
static Properties props;
static StanfordCoreNLP pipeline;
public void initialize(String path)
{
// creates a StanfordCoreNLP object, with POS tagging, lemmatization, NER, parsing, and sentiment
props = new Properties();
props.setProperty("parse.model", path+"edu\\stanford\\nlp\\models\\lexparser\\englishPCFG.ser.gz");
props.setProperty("sentiment.model", path+"edu\\stanford\\nlp\\models\\sentiment\\sentiment.ser.gz");
props.setProperty("annotators", "tokenize, ssplit, parse, sentiment");
pipeline = new StanfordCoreNLP(props);
//LexicalizedParser lp = LexicalizedParser.loadModel("edu/stanford/nlp/models/lexparser/englishPCFG.ser.gz");
}
public SentimentResult getSentimentResult(String text) {
SentimentResult sentimentResult = new SentimentResult();
SentimentClassification sentimentClass = new SentimentClassification();
if (text ! = null && text.length() > 0) {
// run all Annotators on the text
Annotation annotation = pipeline.process(text);
for (CoreMap sentence: annotation.get(CoreAnnotations.SentencesAnnotation.class)) {
// this is the parse tree of the current sentence
Tree tree = sentence.get(SentimentCoreAnnotations.SentimentAnnotatedTree.class);
SimpleMatrix sm = RNNCoreAnnotations.getPredictions(tree);
String sentimentType = sentence.get(SentimentCoreAnnotations.SentimentClass.class);
sentimentClass.setVeryPositive((double)Math.round(sm.get(4) *100d));
sentimentClass.setPositive((double)Math.round(sm.get(3) *100d));
sentimentClass.setNeutral((double)Math.round(sm.get(2) *100d));
sentimentClass.setNegative((double)Math.round(sm.get(1) *100d));
sentimentClass.setVeryNegative((double)Math.round(sm.get(0) *100d));
sentimentResult.setSentimentScore(RNNCoreAnnotations.getPredictedClass(tree));
sentimentResult.setSentimentType(sentimentType);
sentimentResult.setSentimentClass(sentimentClass);
}
}
Return sentimentResult;
}
}
```

Table 6: MAE and accuracy for emotional intelligence.

No. of responses	MAE	Accuracy
200	6.67	95
400	8.67	96.75
600	12	97
800	14.67	97.25
1000	16.67	97.5

Table 7: Eysenck personality questionnaire results.

Eysenck personality			
Standard evaluation		Proposed evaluation	
Result	Count	Result	Count
Psychoticism	35	Psychoticism	35
		Extroversion	0
		Neuroticism	0
Extroversion	665	Psychoticism	0
		Extroversion	665
		Neuroticism	0
Neuroticism*	300	Psychoticism	0
		Extroversion	0
		Neuroticism	300

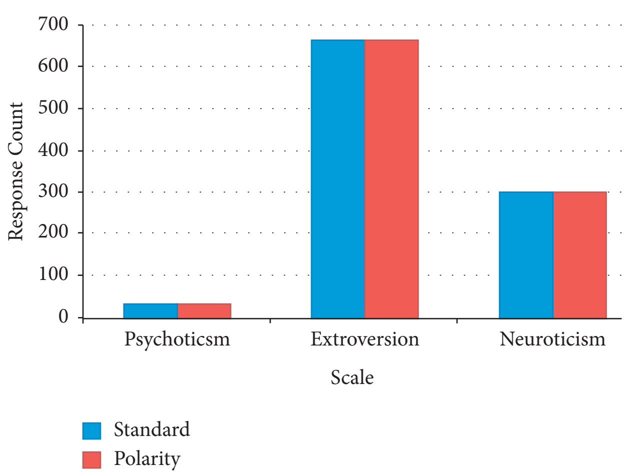


Figure 6: Scale value for EP (standard vs. polarity).

Table 8: MAE and accuracy for Eysenck personality.

No of responses	MAE	Accuracy
200	0	100
400	0	100
600	0	100
800	0	100
1,000	0	100

standard and polarity evaluation produce the same result for all the scale values.

Figure 10 shows the PANAS questionnaire scale value: positive, negative, and neutral.

Table 9: Self-determination scale result.

Self-determination scale			
Standard evaluation		Proposed evaluation	
Result	Count	Result	Count
Low	240	Low	215
		High	25
High	760	Low	20
		High	740

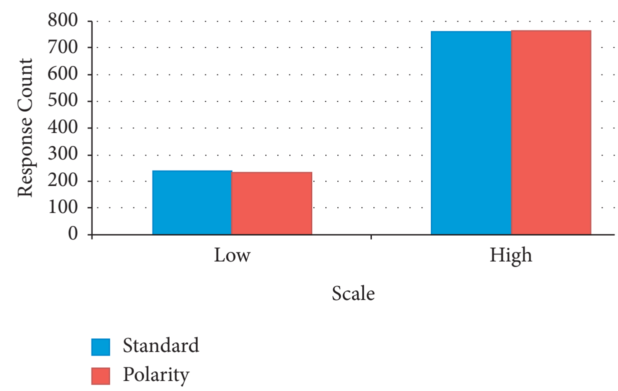


Figure 7: Scale value for SDS (standard vs. polarity).

Table 10: MAE and accuracy self-determination scale.

No. of responses	MAE	Accuracy
200	5	95.5
400	10	96.25
600	10	95
800	5	96.25
1,000	2	94

Table 11: MAE and accuracy for general self-efficacy.

No. of responses	MAE	Accuracy
200	5	97.5
400	15	96.25
600	20	96.67
800	25	96.91
1,000	25	97.5

Table 12: Self-efficacy result.

General self-efficacy			
Standard evaluation		Proposed evaluation	
Result	Count	Result	Count
Low	30	Low	30
		High	0
High	970	Low	25
		High	945

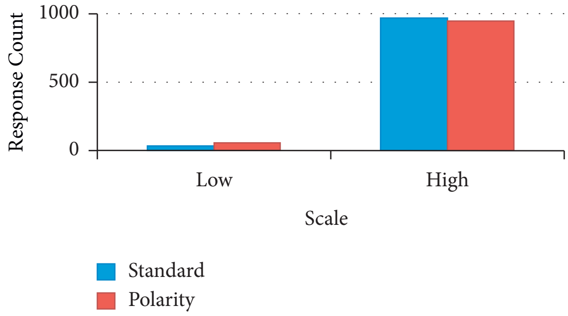


Figure 8: Scale value for GSE (standard vs. polarity).

Table 13: Rosenberg's self-esteem result.

Rosenberg's self-esteem			
Standard evaluation		Proposed evaluation	
Result	Count	Result	Count
Low	0	Low	0
		Normal	0
		High	0
Normal	145	Low	15
		Normal	120
		High	10
High	855	Low	15
		Normal	20
		High	820

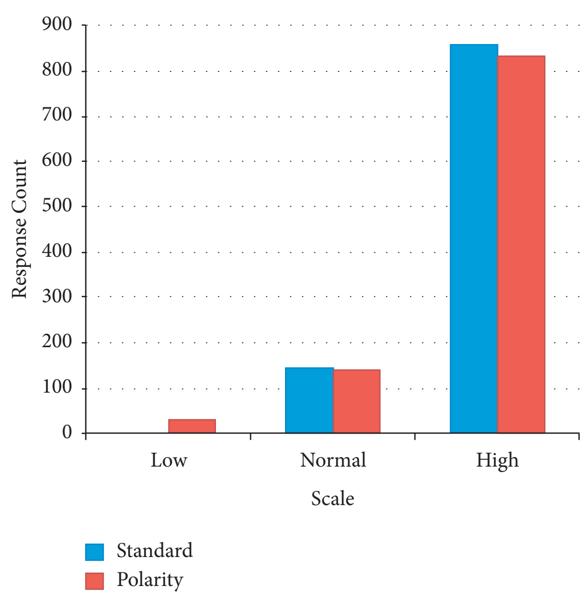


Figure 9: Scale value for RSE.

Table 14: MAE and accuracy for Rosenberg's self-esteem.

No. of responses	MAE	Accuracy
200	5.3	91
400	10	92.5
600	13.3	92.5
800	16.67	93.12
1,000	20	94

Table 15: Positive and negative affect schedule result.

Positive and negative affect schedule			
Standard evaluation		Proposed evaluation	
Result	Count	Result	Count
Positive	805	Positive	805
		Negative	0
		Neutral	0
Negative	125	Positive	0
		Negative	125
		Neutral	0
Neutral	70	Positive	0
		Negative	0
		Neutral	70

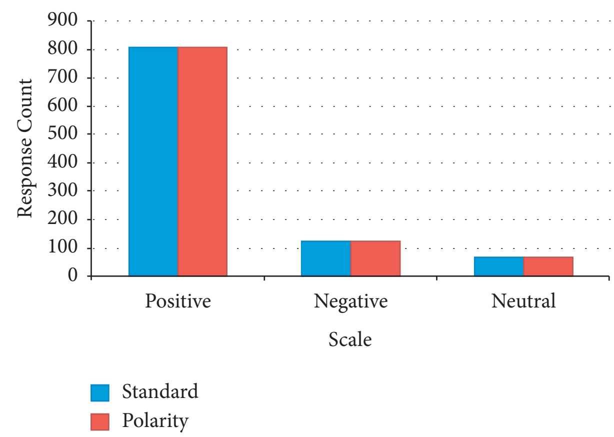


Figure 10: Scale value for PANAS (standard vs. polarity).

Table 16: MAE and accuracy for positive and negative affect schedule.

No. of responses	MAE	Accuracy
200	0	100
400	0	100
600	0	100
800	0	100
1,000	0	100

Table 17: Oxford happiness result.

Oxford happiness			
Standard evaluation		Proposed evaluation	
Result	Count	Result	Count
Happy	250	Happy	200
		Moderately happy	35
		Unhappy	15
Moderately happy	745	Happy	50
		Moderately happy	690
		Unhappy	5
Unhappy	5	Happy	0
		Moderately happy	0
		Unhappy	5

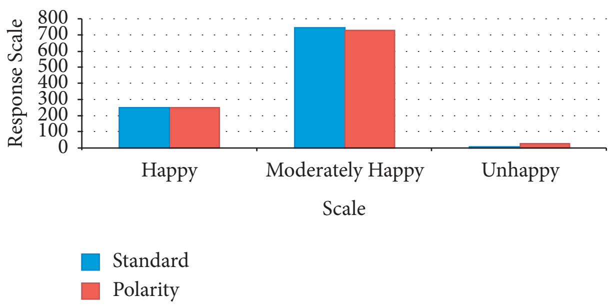


Figure 11: Scale value Oxford happiness (standard vs. polarity).

Table 18: MAE and accuracy for Oxford happiness.

No. of responses	MAE	Accuracy
200	3.35	88
400	6.67	87.5
600	6.67	87.33
800	13.34	87.5
1,000	13.34	89.5

Table 16 shows the MAE and accuracy comparison of PANAS for the different numbers of responses. The result produces zero error and 100% accuracy for all different numbers of responses.

Table 17 shows the Oxford happiness questionnaire standard and proposed polarity-based result. Figure 11 shows the OH questionnaire scale value happy, moderately, happy, and unhappy. Table 18 gives MAE and the accuracy of the Oxford happiness questionnaire.

6. Conclusion

The task of sentiment analysis for questionnaire data was the focus of this study. The main goal was to develop a mechanism for analyzing questions and students' emotions based on closed-ended responses. Quest SA is a tool for assessing questionnaire sentiments and students' emotions proposed in this paper. The students' replies are gathered using a closed-ended questionnaire, and the students' emotions are identified using polarity-based sentiment analysis in this study. The performance of the study task is evaluated using seven real-time surveys (emotional intelligence, Eysenck personality, self-determination scale, general self-efficacy, Rosenberg's self-esteem, positive and negative affect schedule, and Oxford happiness). The suggested Quest_SA accurately predicts students' emotions compared to established questionnaire evaluation methods. The proposed system's accuracy is comparable to that of the traditional method. When opposed to traditional evaluation, categorizing the result is simple. Because the traditional evaluation with range of values takes long time than the proposed evaluation with polarity score, multimodal SA techniques are probably going to be in high demand in the near future.

Table 15 shows the MAE and OH accuracy comparison for a different number of responses. Again, the results proved that the proposed system works similarly to the traditional system with good accuracy.

Data Availability

The data that support the findings of this study are not available in any public repository.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, 2013.

[2] Z. Kastrati, F. Dalipi, A. S. Imran, K. Pireva Nuci, and M. A. Wani, "Sentiment analysis of students' feedback with NLP and deep learning: a systematic mapping study," *Applied Sciences*, vol. 11, no. 9, p. 3986, 2021.

[3] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning," *IEEE Access*, vol. 8, Article ID 23522, 2020.

[4] A. E. O. Carosia, G. P. Coelho, and A. E. A. Silva, "Analyzing the Brazilian financial market through Portuguese sentiment analysis in social media," *Applied Artificial Intelligence*, vol. 34, no. 1, pp. 1–19, 2020.

[5] N. Capuano, L. Greco, P. Ritrovato, and M. Vento, "Sentiment analysis for customer relationship management: an incremental learning approach," *Applied Intelligence*, vol. 51, no. 6, pp. 3339–3352, 2020.

[6] S. K. Sharma, M. Daga, and B. Gemini, "Twitter sentiment analysis for brand reputation of smart phone companies in India," in *Proceedings of ICETIT 2019*, pp. 841–852, Springer, Heidelberg, Germany, 2020.

[7] A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, "Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets," *IEEE Access*, vol. 8, Article ID 181074, 2020.

[8] P. Chauhan, N. Sharma, and G. Sikka, "The emergence of social media data and sentiment analysis in election prediction," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, pp. 1–27, 2020.

[9] J. Zhou and J. M. Ye, "Sentiment analysis in education research: a review of journal publications," *Interactive Learning Environments*, pp. 1–13, 2020.

[10] D. Yang, R. Kraut, and C. P. Rosé, "Exploring the effect of student confusion in massive open online courses," *Journal of Educational Data Mining*, vol. 8, no. 1, pp. 52–83, 2016.

[11] I. Sindhu, S. M. Daudpota, K. Badar, M. Bakhtyar, J. Baber, and M. Nurunnabi, "Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation," *IEEE Access*, vol. 7, Article ID 108729, 2019.

[12] M. V. Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis-A review of research topics, venues, and top cited papers," *Computer Science Review*, vol. 27, pp. 16–32, 2018.

[13] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016.

[14] S. Sun, C. Luo, and J. Chen, "A review of natural language processing techniques for opinion mining systems," *Information Fusion*, vol. 36, pp. 10–25, 2017.

[15] Q. Yang, Y. Rao, H. Xie, J. Wang, F. L. Wang, and W. H. Chan, "Segment-level joint topic-sentiment model for online review

analysis," *IEEE Intelligent Systems*, vol. 34, no. 1, pp. 43–50, 2019.

[16] S. Zhang, Z. Wei, Y. Wang, and T. Liao, "Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary," *Future Generation Computer Systems*, vol. 81, pp. 395–403, 2018.

[17] N. Mukhtar, M. A. Khan, and N. Chiragh, "Lexicon-based approach outperforms supervised machine learning approach for Urdu sentiment analysis in multiple domains," *Telematics and Informatics*, vol. 35, no. 8, pp. 2173–2183, 2018.

[18] C. Wan, Y. Peng, K. Xiao, X. Liu, T. Jiang, and D. Liu, "An association-constrained LDA model for joint extraction of product aspects and opinions," *Information Sciences*, vol. 519, pp. 243–259, 2020.

[19] Z.-P. Fan, Y.-J. Che, and Z.-Y. Chen, "Product sales forecasting using online reviews and historical sales data: a method combining the bass model and sentiment analysis," *Journal of Business Research*, vol. 74, pp. 90–100, 2017.

[20] G. A. Ruz, P. A. Henríquez, and A. Mascareño, "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers," *Future Generation Computer Systems*, vol. 106, pp. 92–104, 2020.

[21] N. Majumder, S. Poria, H. Peng, N. Chhaya, E. Cambria, and A. Gelbukh, "Sentiment and sarcasm classification with multitask learning," *IEEE Intelligent Systems*, vol. 34, no. 3, pp. 38–43, 2019.

[22] M. Atzeni, A. Dridi, and D. Reforgiato Recupero, "Using frame-based resources for sentiment analysis within the financial domain," *Progress in Artificial Intelligence*, vol. 7, no. 4, pp. 273–294, 2018.

[23] P. Kaewyong, A. Sukprasert, N. Salim, and F. A. Phang, "The possibility of students' comments automatic interpret using lexicon based sentiment analysis to teacher evaluation," in *Proceedings of the 3rd International Conference on Artificial Intelligence and Computer Science 2015 (AICS2015)*, pp. 179–189, Penang, Malaysia, September 2015.

[24] Q. Rajput, S. Haider, and S. Ghani, "Lexicon-based sentiment analysis of teachers' evaluation," *Applied Computational Intelligence and Soft Computing*, vol. 2016, Article ID 2385429, 12 pages, 2016.

[25] Z. Nasim, Q. Rajput, and S. Haider, "Sentiment analysis of student feedback using machine learning and lexicon based approaches," in *Proceedings of the 2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*, pp. 1–6, Langkawi, Malaysia, July 2017.

[26] K. Z. Aung and N. N. Myo, "Sentiment Analysis of Students Comment Using Lexicon Based Approach," in *Proceedings of the 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pp. 149–154, Wuhan, China, May 2017.

[27] M. Z. Asghar, F. Subhan, H. Ahmad et al., "Senti-eSystem : a sentiment-based eSystem -using hybridized fuzzy and deep neural network for measuring customer satisfaction," *Software: Practice and Experience*, vol. 51, no. 3, pp. 571–594, 2021.

[28] T. Gadekallu, A. Soni, D. Sarkar, and L. Kuruva, "Application of sentiment analysis in movie reviews," *Advances in Business Information Systems and Analytics-Sentiment Analysis and Knowledge Discovery in Contemporary Business*, pp. 77–90, 2019.

[29] L. K. Ramasamy, S. Kadry, and S. Lim, "Selection of optimal hyper-parameter values of support vector machine for sentiment analysis tasks using nature-inspired optimization methods," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 290–298, 2021.

[30] F. Khan, R. L. Kumar, and S. Kadry, "Hybrid reality-based education expansion system for non-traditional learning," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 7, no. 1, pp. 185–192, 2021.

[31] M. A. Jayanthi, R. L. Kumar, and S. Swathi, "Investigation on association of self-esteem and students' performance in academics," *International Journal of Grid and Utility Computing*, vol. 9, no. 3, pp. 211–219, 2018.

[32] M. A. Jayanthi, R. L. Kumar, A. Surendran, and K. Prathap, "Research contemplate on educational data mining," in *Proceedings of the 2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, pp. 110–114, Coimbatore, India, October 2016.

[33] P. Hills and M. Argyle, "The Oxford Happiness Questionnaire: a compact scale for the measurement of psychological well-being," *Personality and Individual Differences*, vol. 33, no. 7, pp. 1073–1082, 2002.

[34] D. Goleman, *Working with Emotional Intelligence*, Bantam, New York, 1998.

[35] S. B. G. Eysenck, H. J. Eysenck, and P. Barrett, "A revised version of the psychoticism scale," *Personality and Individual Differences*, vol. 6, no. 1, pp. 21–29, 1985.

[36] A. Abdul Aziz and A. Starkey, "Predicting supervise machine learning performances for sentiment analysis using contextual-based approaches," *IEEE Access*, vol. 8, Article ID 17722, 2020.

[37] S. Poria, N. Majumder, R. Mihalcea, and E. Hovy, "Emotion recognition in conversation: research challenges, datasets, and recent advances," *IEEE Access*, vol. 7, Article ID 100943, 2019.

Indonesian Journal of Electrical Engineering and Computer Science
Vol. 29, No. 1, January 2023, pp. 545~552
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v29.i1.pp545-552

Reclust: an efficient clustering algorithm for mixed data based on reclustering and cluster validation

Amala Jayanthi Maria Soosai Arockiam[1], Elizabeth Shanthi Irudhayaraj[2]
[1]Department of Computer Applications, Kumaraguru College of Technology, Coimbatore, India
[2]Department of Computer Science, Avinashilingam Institute of Home Science and Higher Studies, Coimbatore, India

Article Info

Article history:

Received Apr 8, 2022
Revised Sep 15, 2022
Accepted Oct 12, 2022

Keywords:

Clustering
Hierarchical cluster
K-means
Mixed data
Numerical and categorical data
Re-clustering

ABSTRACT

Clustering is a significant approach in data mining, which seeks to find groups or clusters of data. Both numeric and categorical features are frequently used to define the data in real-world applications. Several different clustering algorithms are proposed for the numerical and categorical datasets. In clustering algorithms, the quality of clustering results is evaluated using cluster validation. This paper proposes an efficient clustering algorithm for mixed numerical and categorical data using re-clustering and cluster validation. Initially, the mixed dataset is clustered with four traditional clustering algorithms like expectation-maximization (EM), hierarchical cluster (HC), k-means (KM), and self-organizing map (SOM). These four algorithms are validated, and the best algorithm is selected for re-clustering. It is an iterative process for improving the quality of cluster results. The incorrectly clustered data is iteratively re-clustered and evaluated based on the cluster validation. The performance of the proposed clustering method is evaluated with a real-time dataset in terms of purity, normalized mutual information, rand index, precision, and recall. The experimental results have shown that the proposed reclust algorithm achieves better performance compared to other clustering algorithms.

Corresponding Author:

Amala Jayanthi Maria Soosai Arockiam
Department of Computer Science, Kumaraguru College of Technology
Coimbatore, Tamil Nadu, India
Email: research.amala@gmail.com

1. INTRODUCTION

Clustering analysis is one of the most important approaches in data mining, and it seeks to determine the nature of groupings or clusters of data objects in attributes space. Clustering methods are employed in a variety of applications [1], including social network analysis [2], knowledge discovery, image processing, text and sentiment analysis [3]. Clustering analysis seeks to group data objects with similar properties together, and those with distinct characteristics into separate clusters. Hierarchical and partitional clustering methods are the two types of clustering algorithms [4]. Data are dispersed into a dendrogram of layered segments using a split or agglomerative technique in hierarchical clustering algorithms. Data are partitioned into a certain number of clusters by minimizing an objective cost function in partitional clustering algorithms.

For specific kinds of information, clustering algorithms have been developed. Continuous values are used to represent numerical data, whereas categorical data, which is a subset of discrete data, can only have a finite number of values. Many real-world applications use categorical data, such as name, gender, and educational level. Both numerical and category values were present in the mixed datasets. Real-world data is frequently of various sorts. Medical data, for example, includes categorical and numerical values such as age, height, weight, and salary, as well as categorical and numerical values such as nationality, gender, employment, education [5], marital status, and chest pain type [6]. When a dataset comprises both numerical and categorical variables, the issue of determining the similarity of two data becomes more complicated [7]. Splitting the numeric and categorical elements of a mixed dataset and finding the Euclidean distance between two data points for numeric characteristics and the Hamming distance for categorical features is a simple technique for solving the similarity problem [8].

For clustering mixed data, several techniques have been developed. To cluster heterogeneous data, Huang [9] presented the well-known k-prototypes technique, which merged the k-means and k-modes approaches. The k-prototypes algorithm was improved in [10] by incorporating attribute influence and enhancing the cluster center representation. The unsupervised feature learning (UFL) approach was developed by Lam *et al*. [11] by combining the fuzzy adaptive resonance theory (ART) with the UFL. The approach Kay-means for mixed large data sets (KAMILA) introduced by Foss *et al*. [12] can directly deal with multiple types of attributes and requires fewer parameters. Chen and He [13] used the principle of density clustering to present a self-adaptive peak density clustering technique. Most mixed data clustering algorithms have two main goals: to develop new approaches to construct novel measures of similarity between mixed characteristics and to cluster data using previous or new strategies to obtain a local optimum result.

This paper proposes an efficient clustering algorithm for mixed numerical and categorical data based on re-clustering and cluster validation called reclust. The proposed method contains three important processes: initial clustering, validation, and re-clustering. The initial clustering process uses four traditional clustering algorithms such as expectation-maximization (EM), hierarchical cluster (HC), k-means (KM), and self-organizing map (SOM). The validation process evaluates the clustering result. The re-clustering process re-clusters the incorrectly clustered data. The validation and the re-clustering process is an iterative process [14]. It improves the quality of cluster results.

The remaining part of this research paper is as follows: section 2 describes the research background including different clustering methods for numerical and categorical data and also explains clustering algorithms used in research. Then, the proposed methodology is explained in section 3. The performance of the proposed work is analyzed in section 4-the conclusion and the future work of this research work are provided in section 5.

2. RESEARCH BACKGROUND

2.1. Mixed data clustering

Clustering mixed data is a difficult process that is rarely accomplished using well-known clustering algorithms developed for a certain type of data. It is common knowledge that converting one type to another is insufficient since it may result in data loss [15]. For clustering mixed datasets, Que *et al*. [16] suggest a similarity measurement using entropy-based weighting. An automatic categorization technique is used to convert numerical data into category data. The relevance of various attributes is then denoted using an entropy-based weighting technique.

Li *et al*. [17] offer a mixed data clustering technique with a noise-filtered distribution centroid and an iterative weight modification strategy. It defines a noise-filtered distribution centroid for categorical attributes. By integrating the mean and noise-filtered distribution centroid, this method displays the cluster centre with mixed properties. The frequency of occurrences for each potential value of the categorical attributes in a cluster is more accurately recorded by the noise-filtered distribution centroid.

Jia and Cheung [18] show how to cluster data using soft subspace clustering with both numerical and categorical features. The model is based on the definition of object-cluster similarity and is attribute-weighted. Using a uniform weighting approach for numerical and categorical qualities, the attribute-to-cluster contribution is measured by accounting for both inter-cluster difference and intra-cluster similarity.

For data with heterogeneous features, D'Urso and Massari [19] suggest a fuzzy clustering model. Different sorts of variables, or qualities, can be considered using the clustering model. This result is obtained by using a weighting system to combine the dissimilarity measurements for each attribute, yielding a distance measure for several attributes. During the optimization phase, the weights are computed objectively. The weights in the clustering findings represent the importance of each attribute type. Rodriguez *et al*. [20] suggest a multipartition clustering process that combines Bayesian network factorization and the variational Bayes framework to efficiently handle mixed data.

2.2. K-means clustering algorithm

Let X=x 1,x 2,...,x n be a data collection in a d-dimensional Euclidean space Rd, and A=a 1,a 2,...,a c be the c cluster centres, with d ik=x i-a k as its euclidean norm. Let U= ik _(nc), where _ik is a binary variable (i.e., _ik 0,1) that indicates whether the data point xi belongs to the kth cluster, k=1,2,...,c. By minimizing the k-means objective function, the k-means clustering method is iterated via the updating equations for cluster centres and memberships [12]: $J(U,A)=\sum_{i=1}^{n}\sum_{k=1}^{c}\mu_{ik}\|x_i-a_k\|^2$ as $a_k=\sum_{i=1}^{n}\mu_{ik}x_{ij}/\sum_{i=1}^{n}\mu_{ik}$ and $\mu_{ik}=\begin{cases}1 \text{ if } \|x_i-a_k\|^2-\min\limits_{1\le k\le c}\|x_i-a_k\|^2\\ \quad 0 \text{ Otherwise}\end{cases}$

2.3. Hierarchical clustering

In Algorithm 1 describe the hierarchical clustering pseudocode. Methods that use hierarchical clustering build a hierarchy of clusters that are arranged from top to bottom (or bottom to up). The hierarchical algorithms require both of the following to build clusters:

- Similarity matrix–this is created by determining how similar each pair of mixed data values are. The shape of the clusters is influenced by the similarity measure used to generate the similarity matrix.
- Linkage criterion–this establishes the distance between sets of observations as a function of pairwise distances.

Algorithm 1. Hierarchical clustering pseudocode
```
C = {Ci – {xi} | xi ϵ D}
Δ = {δ(xi, xj); xi, xj ϵ D}
Repeat
 Find the closest pair of clusters Ci, Cj ϵ C
 Cij = Ci U Cj
 C = C \ {{ Ci} U {Cj}} U {Cij}
 Update distance matrix Δ to reflect new clustering
Until |C| = k
```

2.4. Expectation maximization

The EM algorithm in Algorithm 2 finds maximum likelihood parameter estimates in probabilistic models. The iterative technique of expectation maximisation (EM) alternates between two steps: expectation (E) and maximum (M). To cluster data, EM employs the finite Gaussian mixtures model, which iteratively estimates a set of parameters until the desired convergence value is obtained. Each of the K probability distributions in the mixture corresponds to a single cluster. A membership probability is assigned to each instance by each cluster [21].

Algorithm 2. EM clustering pseudocode

1. Initialize estimates for $\theta \coloneqq \pi,\mu_1,\sigma_1,\mu_2,\sigma_2$
2. (Expectation) Compute the responsibilities for each data point

$$\gamma_i=\frac{\pi\varphi(x_i;\mu_2,\sigma_2)}{(1-\pi)\varphi(x_i;\mu_1,\sigma_1)+\pi\varphi(x_i;\mu_2,\sigma_2)}$$

3. (Maximization) Update the estimates for the parameters using the maximum likelihood estimator formula. All sums are taken across the data indexed by i and are just means/standard deviations weighted by the responsibilities γ

$$\mu_2=\frac{\sum\gamma_i x_i}{\sum\gamma_i}\,\sigma_2=\frac{\sum\gamma_i(x_i-\mu_2)^2}{\sum\gamma_i}\,\pi=\frac{1}{n}\sum\gamma_i$$

4. Repeat steps 2 and 3 until the parameters converge to a local optimum.

2.5. Self organization map

The SOM algorithm in Algorithm 3 is a classic unsupervised learning neural network model that clusters input data with similarities. It employs an unsupervised learning methodology and used a competitive learning algorithm to train its network. In order to minimise complex issues for straightforward interpretation, SOM is utilised for clustering and mapping (or dimensionality reduction) procedures to map multidimensional data onto lower-dimensional spaces. The input layer and the output layer are the two layers that make up SOM. The SOM merges the clustering and projection operations (reduce the dimensionality of information).

Algorithm 3. Self organization map

1. Initialize the weight wj, neighborhood parameter Np, k = 0, and learning rate μ=1.0;
2. Select random vector x from input data
3. Compute and select the winning neuron i.e Best Matching Unit based on a distance measure and neighborhood function. The empirical index of the winning neurons is determined as follows:

$$i(x)=\arg\min_{1\le j\le d}\|x-w_j\|$$

4. Update the weight vector of winning neurons

$$w_i(k+1)=\begin{cases}w_i(k)+\mu[x(k)-w_i(k)]\ i\in N_p(k)\\ w_i(k)\ i\notin N_p(k)\end{cases}$$

5. Update the parameters
6. Repeat Steps 2, 3, and 4 until the stopping criteria are met.

3. PROPOSED METHOD

This section explains the proposed clustering algorithm for mixed numerical and categorical data [22] based on re-clustering and cluster validation called reclust. The proposed method contains three important processes: initial clustering, validation, and re-clustering. The initial clustering process uses four traditional clustering algorithms such as EM, HC, KM, and SOM. The validation process evaluates the clustering result. The re-clustering process re-clusters the incorrectly clustered data. The validation and the re-clustering process is an iterative process. It improves the quality of cluster results.

Let D be the mixed dataset consisting of n instances, indicates as {d1, d2, …, dn}. The dataset D has ac categorical attributes and au numerical attributes. Then $d_i (1 \leq i \leq n)$ can be denoted as $[d_i^c, d_i^u]$ with $d_i^c = [d_{i1}^c, d_{i2}^c, \dots, d_{i,a_c}^c]$ and $d_i^u = [d_{i1}^u, d_{i2}^u, \dots, d_{i,a_u}^u]$. Cluster the dataset D into k clusters C= {C1, C2, …., Ck}. $C_i \cap C_j = \emptyset$, $\bigcup_{i=1}^{k} C_i = C (i, j = 1,2, \dots, k, i \neq j)$. Algorithm 4 explains the reclust clustering algorithm.

Algorithm 4. Reclust
```
Input: Dataset D = { d1, d2, …, dn}, Number of Cluster k
Output: Clustering Result
1. Initial Clustering
 1a. EMcls = Apply EM(D, k)
 1b. HCcls = Apply HC(D, k)
 1c. KMcls = Apply KM(D, k)
 1d. SOMcls = Apply SOM(D, k)
2. Cluster Validation
 2a. EMeval = evaluateCluster(EMcls, D)
 2b. HCeval = evaluateCluster(HCcls, D)
 2c. KMeval = evaluateCluster(KMcls, D)
 2d. SOMcls = evaluateCluster(SOMcls, D)
 2e. minCls = Min (EMeval, HCeval, KMeval, SOMeval)
 2f. incorrectD = incorrectlyClusteredData(D)
3. Reclustering
 3a. While (the stop criterion is not met)
 3b. recls= Cluster incorrectD using minCls
 3c. reclsEval = evaluateCluster (recls)
 3d. subD = incorrectlyClusteredData(incorrectD)
 3e. incorrect =subD
 3f. End While
```

In this algorithm, step 1 applies four traditional clustering algorithms. Step 2 evaluates the cluster results. The evaluateCluster uses classes to cluster evaluation method. It builds clustering after ignoring the class attribute. It then allocates classes to the clusters during the test phase, depending on the majority value of the class feature within each cluster. The classification error is then calculated based on this assignment. Step 2e finds the minimum error value of four traditional clustering algorithms. Step 2f extracts the incorrectly clustered data from the evaluation results. Step 3 is an iterative re-clustering, which clusters the incorrect data and evaluates the clustering result. The stop criterion for the re-clustering step is either a minimum error value or a minimum number of instances in incorrect clustered data.

4. EXPERIMENTAL RESULT

This section evaluates the performance of the proposed work through experiments. Three publicly available data sets and students' data with seven questionnaires are used to analyze the cluster results. Table 1 shows the summary of the dataset used for experiments. The following metrics are used to evaluate the clustering results: rand index (RI), precision (Pre), and recall (Rec). These evaluation metrics are computed using the classes to cluster assignment (CCA) table shown in Table 2.

Let D={D1, D2, D3,…, Dn} be the dataset contains n number of instances, C={C1, C2, …,Ck} denotes set of k clusters generated from D using clustering algorithm and P = {P1, P2, …, Pc}denotes set of c true classes of D. In table 2, aij represents the number of common instances between Pi and Cj i.e aij = |Pi ∩ Cj|. SPi and SCj denote the number of instances in Pi and Cj.

The evaluation metrics are computed as shown in:

$$Purity = \frac{1}{n} \sum_k \max_c |a_{kc}|$$

$$ARI = \frac{\sum_{ij}\binom{a_{ij}}{2} - \left[\sum_i \binom{SP_i}{2} \sum_j \binom{SC_j}{2}\right] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{SP_i}{2} + \sum_j \binom{SC_j}{2}\right] - \left[\sum_i \binom{SP_i}{2} \sum_j \binom{SC_j}{2}\right] / \binom{n}{2}}$$

here $\binom{n}{2} = n(n-1)/2$,

$$NMI = \frac{\sum_{i=1}^{c}\sum_{j=1}^{k} a_{ij}\log\left(\frac{a_{ij}*n}{SP_i*SC_j}\right)}{\sqrt{\sum_{i=1}^{c} SP_i \log\left(\frac{SP_i}{n}\right)\sum_{j=1}^{k} SC_j log\left(\frac{SC_j}{n}\right)}}$$

$$Pre = \frac{1}{c}\sum_{i=1}^{c}\frac{\max_k a_{ki}}{SP_i}S$$

$$Rec = \frac{1}{k}\sum_{i=1}^{k}\frac{\max_c a_{ci}}{SC_i}$$

In this experiment, the number of clusters to be found was equal to the number of classes in the data set i.e., c = k. Larger values of RI, Pre, and Rec indicate better clustering results. Table 3 shows the Classes for Cluster Assignment for the emotional intelligence dataset. Most of the classes are correctly clustered.

Tables 4-9 shows CCA for EPQ, GSE, EHQ, PNA, RSE [23], SDS datasets. Table 10 shows the comparison of evaluation metrics for different datasets. The metrics RI, Precision, and Recall is compared with ABC-K-Prototypes [24], CCS-K-Prototypes [1], and Multi-view K-Prototype [25]. Table 11 and Figure 1 shows the Rand Index comparison. Table 12 and Figure 2 depict the precision comparison. Table 13 and Figure 3 depicts the recall comparison.

Table 1. Dataset summary

Dataset Type	Dataset	# Instances	# Numerical Features	# Categorical Features	# Classes
Student Info with Question Response	Emotional Intelligence (EIQ)	1000	2	11	3
	Eysenck Personality (EPQ)	1000	2	11	3
	General Self Efficacy (GSE)	1000	2	11	2
	Emotional Happiness (EHQ)	1000	2	11	3
	Positive /Negative Attitude (PNA)	1000	2	11	3
	Self Esteem(RSE)	1000	2	11	3
	Self Determination (SDS)	1000	2	11	2
Medical	Heart	293	7	6	5
	Dermatology	358	1	33	6
Credit Card	Credit	653	6	9	2

Table 2. Classes to cluster assignment table

	C_1	C_2	….	C_k	Sum
P_1	a_{11}	a_{12}	….	a_{1k}	SP_1
P_2	a_{21}	a_{22}	….	a_{2k}	SP_2
⋮	⋮	⋮	⋮	⋮	⋮
P_c	a_{c1}	a_{c2}	….	a_{ck}	SP_c
Sum	SC_1	SC_2	….	SC_k	

Table 3. CCA for emotional intelligence dataset

CCA	Assigned Cluster			
		High	Average	Low
Actual Classes	High	700	0	0
	Average	15	265	10
	Low	8	0	2

Table 4. CCA for eysenck personality dataset

CCA	Assigned Cluster			
		Extroversion	Psychoticism	Neuroticism
Actual Classes	Extroversion	665	0	0
	Psychoticsm	7	25	3
	Neuroticism	20	0	280

Table 5. CCA for self efficacy

CCA	Assigned Cluster		
		High	Low
Actual Classes	High	960	10
	Low	4	26

Table 6. CCA for emotional happiness

CCA		Assigned Cluster		
		Happy	Moderately_happy	Unhappy
Actual Classes	Happy	238	12	0
	Moderately_happy	39	692	0
	Unhappy	0	4	15

Table 7. CCA for positive/negative attitude

Table 8. CCA for self esteem

CCA		Assigned Cluster		
		Positive	Negative	Neutral
Actual Classes	Positive	756	35	14
	Negative	12	113	0
	Neutral	0	4	66

CCA		Assigned Cluster		
		High	Normal	Low
Actual Classes	High	890	0	16
	Normal	6	52	0
	Low	6	0	30

Table 9. CCA for self determination

CCA		Assigned Cluster	
		High	Low
Actual Classes	High	745	15
	Low	20	220

Table 10. Evaluation metrics comparison

Data Set	EIQ	EPQ	GSE	EHQ	PNA	RSE	SDS	Heart	Dermatology	Credit Card
Purity	0.975	0.97	0.986	0.945	0.935	0.972	0.965	0.724	0.911	0.928
RI	0.908	0.901	0.821	0.798	0.774	0.838	0.859	0.899	0.865	0.947
NMI	0.823	0.83	0.606	0.689	0.645	0.715	0.735	0.817	0.886	0.9
Pre	0.905	0.883	0.928	0.896	0.929	0.904	0.948	0.777	0.919	0.983
Rec	0.934	0.983	0.859	0.946	0.851	0.88	0.955	0.684	0.936	0.986

Table 11. RI comparison

Dataset	ABC-K	CCS-K	Multi-View	Reclust
Heart	0.667	0.680	0.684	0.899
Dermatology	0.689	0.694	0.691	0.865
Credit Card	0.673	0.674	0.695	0.947

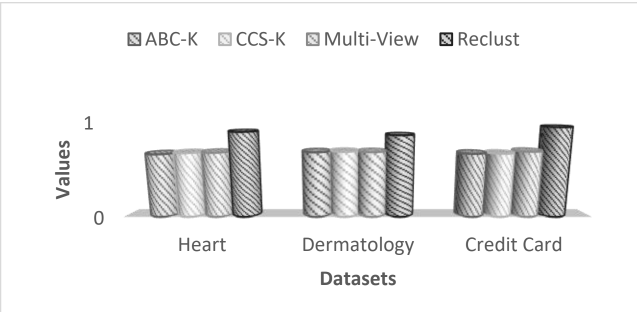


Figure 1. Rand index comparison

Table 12. Precision comparison

Dataset	ABC-K	CCS-K	Multi-View	Reclust
Heart	0.658	0.675	0.637	0.777
Dermatology	0.808	0.812	0.809	0.919
Credit Card	0.792	0.814	0.810	0.983

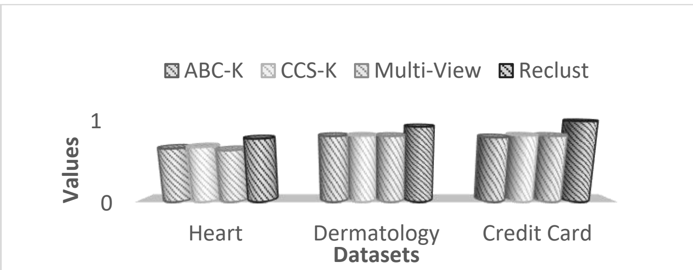


Figure 2. Precision comparison

Table 13. Recall comparison

Dataset	ABC-K	CCS-K	Multi-View	Reclust
Heart	0.379	0.388	0.398	0.684
Dermatology	0.806	0.809	0.807	0.936
Credit Card	0.795	0.796	0.810	0.986

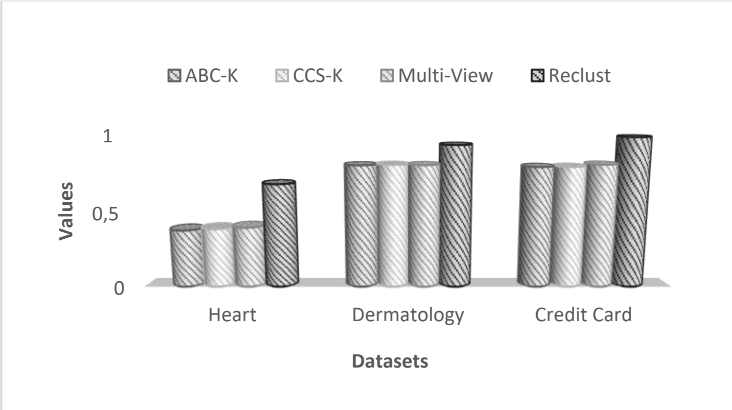


Figure 3. Recall comparison

5. CONCLUSION

Clustering is a typical data mining technique, and clustering mixed datasets into meaningful groups is possible since mixed items are ubiquitous in real-world datasets. This research presents an effective clustering approach for grouping mixed numerical and categorical datasets. Furthermore, iterative re-clustering and cluster validation enhance the clustering results. In terms of clustering purity, NMI, rand index, precision, and recall, the suggested reclust algorithm was tested on several datasets. The results of the experiments confirm the reclust algorithm's superior performance.

REFERENCES

[1] J. Ji, W. Pang, Z. Li, F. He, G. Feng, and X. Zhao, "Clustering mixed numeric and categorical data with cuckoo search," *IEEE Access*, vol. 8, pp. 30988–31003, 2020, doi: 10.1109/ACCESS.2020.2973216.

[2] L. K. Ramasamy, S. Kadry, Y. Nam, and M. N. Meqdad, "Performance analysis of sentiments in Twitter dataset using SVM models," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2275–2284, 2021, doi: 10.11591/ijece.v11i3.pp2275-2284.

[3] Amala Jayanthi M. and E. Shanthi I., "Role of Educational data mining in student learning processes with sentiment analysis," *International Journal of Knowledge and Systems Science*, vol. 11, no. 4, pp. 31–44, Oct. 2020, doi: 10.4018/IJKSS.2020100103.

[4] J. Han, J. Pei, and H. Tong, *Data Mining*, 3rd ed. Elsevier, 2012.

[5] M. A. Jayanthi, R. L. Kumar, A. Surendran, and K. Prathap, "Research contemplate on educational data mining," in *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, Oct. 2016, pp. 110–114, doi: 10.1109/ICACA.2016.7887933.

[6] D.-T. Dinh, V.-N. Huynh, and S. Sriboonchitta, "Clustering mixed numerical and categorical data with missing values," *Information Sciences*, vol. 571, pp. 418–442, Sep. 2021, doi: 10.1016/j.ins.2021.04.076.

[7] L. K. Ramasamy, S. Kadry, and S. Lim, "Selection of optimal hyper-parameter values of support vector machine for sentiment analysis tasks using nature-inspired optimization methods," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 290–298, 2021, doi: 10.11591/eei.v10i1.2098.

[8] Ahmad Amir and Khan Shehroz, "Survey of state-of-the-art mixed data clustering algorithms," *IEEE Access*, vol. 8, pp. 318883–31902, 2020.

[9] Z. Huang, "Clustering large data sets with mixed numeric and categorical values," in *Proceedings of the 1st Pacific-Asia Conference on Knowledge Discovery and Data Mining,(PAKDD)*, 1997, pp. 21–34, [Online]. Available: http://reference.kfupm.edu.sa/content/c/l/clustering_large_data_sets_with_mixed_nu_362883.pdf.

[10] J. Ji, T. Bai, C. Zhou, C. Ma, and Z. Wang, "An improved k-prototypes clustering algorithm for mixed numeric and categorical data," *Neurocomputing*, vol. 120, pp. 590–596, Nov. 2013, doi: 10.1016/j.neucom.2013.04.011.

[11] D. Lam, M. Wei, and D. Wunsch, "Clustering data of mixed categorical and numerical type with unsupervised feature learning," *IEEE Access*, vol. 3, pp. 1605–1613, 2015, doi: 10.1109/ACCESS.2015.2477216.

[12] A. Foss, M. Markatou, B. Ray, and A. Heching, "A semiparametric method for clustering mixed data," *Machine Learning*, vol. 105, no. 3, pp. 419–458, 2016, doi: 10.1007/s10994-016-5575-7.

[13] J. Y. Chen and H. H. He, "A fast density-based data stream clustering algorithm with cluster centers self-determined for mixed data," *Information Sciences*, vol. 345, pp. 271–293, 2016, doi: 10.1016/j.ins.2016.01.071.

[14] A. Murugesan, B. Saminathan, F. Al-Turjman, and R. L. Kumar, "Analysis on homomorphic technique for data security in fog computing," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 9, Sep. 2021, doi: 10.1002/ett.3990.

[15] S. Behzadi, N. S. Müller, C. Plant, and C. Böhm, "Clustering of mixed-type data considering concept hierarchies: problem specification and algorithm," *International Journal of Data Science and Analytics*, vol. 10, no. 3, pp. 233–248, Sep. 2020, doi: 10.1007/s41060-020-00216-2.

[16] X. Que, S. Jiang, J. Yang, and N. An, “A Similarity measurement with entropy-based weighting for clustering mixed numerical and categorical datasets,” *Algorithms*, vol. 14, no. 6, p. 184, Jun. 2021, doi: 10.3390/a14060184.

[17] X. Li, Z. Wu, Z. Zhao, F. Ding, and D. He, “A mixed data clustering algorithm with noise-filtered distribution centroid and iterative weight adjustment strategy,” *Information Sciences*, vol. 577, pp. 697–721, Oct. 2021, doi: 10.1016/j.ins.2021.07.039.

[18] H. Jia and Y. M. Cheung, “Subspace clustering of categorical and numerical data with an unknown number of clusters,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3308–3325, Aug. 2018, doi: 10.1109/TNNLS.2017.2728138.

[19] P. D’Urso and R. Massari, “Fuzzy clustering of mixed data,” *Information Sciences*, vol. 505, pp. 513–534, 2019, doi: 10.1016/j.ins.2019.07.100.

[20] F. Rodriguez-Sanchez, C. Bielza, and P. Larrañaga, “Multipartition clustering of mixed data with Bayesian networks,” *International Journal of Intelligent Systems*, vol. 37, no. 3, pp. 2188–2218, Mar. 2022, doi: 10.1002/int.22770.

[21] K. P. Sinaga and M. S. Yang, “Unsupervised K-means clustering algorithm,” *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[22] X. Jin and J. Han, “Expectation maximization clustering,” in *Encyclopedia of Machine Learning and Data Mining*, Boston, MA: Springer US, 2016, pp. 1–2.

[23] M. Amala Jayanthi, S. Swathi, and R. Lakshmana Kumar, “Investigation on association of self-esteem and students’ performance in academics,” *International Journal of Grid and Utility Computing*, vol. 9, no. 3, p. 211, 2018, doi: 10.1504/IJGUC.2018.10015144.

[24] J. Ji, Y. Chen, G. Feng, X. Zhao, and F. He, “Clustering mixed numeric and categorical data with artificial bee colony strategy,” *Journal of Intelligent and Fuzzy Systems*, vol. 36, no. 2, pp. 1521–1530, 2019, doi: 10.3233/JIFS-18146.

[25] J. Ji, R. Li, W. Pang, F. He, G. Feng, and X. Zhao, “A multi-view clustering algorithm for mixed numeric and categorical data,” *IEEE Access*, vol. 9, pp. 24913–24924, 2021, doi: 10.1109/ACCESS.2021.3057113.

BIOGRAPHIES OF AUTHORS

Amala Jayanthi Maria Soosai Arockiam is an assistant professor at Kumaraguru College of Technology in Coimbatore, where she works in the Department of Computer Applications. She has 9 years of academic experience. She earned her master's degree in computer science from Bharathiyar University's St. Joseph's College (Autonomous) in Tiruchirappalli, Tamil Nadu, India. Kalasalingam University awarded her aMCA degree. Data mining is her current study focus. She has 16 research papers published in reputable publications. She can be contacted at email: research.amala@gmail.com.

Dr. Elizabeth Shanthi Irudhayaraj is an Associate Professor of Computer Science at Coimbatore's Avinashilingam University for Women. She has more than 25 years of teaching experience, as well as ten years of research experience. She has a long list of publications to her name. Data mining, information retrieval, objectoriented data bases, cloud computing, and soft computing are some of her areas of interest. She can be contacted at email: elizabeth_cs@avinuty.ac.in.

Avinashilingam Institute for Home Science and Higher Education for Women
(Deemed to be University Estd. u/s 3 of UGC Act 1956, Category A by MHRD)
Re-accredited with 'A++' Grade by NAAC.CGPA 3.65/4, Category I by UGC
Coimbatore – 641 043, Tamil Nadu, India

PLAGIARISM CHECK REPORT (THESES)

1.	Name of the Research Scholar	Amala Jayanthi M
2.	Roll No. and Year of Registration	19PHCSP001, 2019
3.	Department	Computer Science
4.	Name of the Research Guide	Dr. I. Elizabeth Shanthi
5.	Title of the Thesis / Dissertation	An Integrated Framework for Students' Academic Performance Prediction Based on Affective and Cognitive State using Mixed Numeric and Categorical Data
6.	Similarity Content (%) Identified	**6%**
7.	Software Used	Turnitin
8.	Date of Verification	08-08-2023

Note : The report is excluding 14 Consecutive words, Review of Literature and Quoted Materials.

Checked by :

Jhy.M

For **Information Scientist**

M. malarvizhi

Research Scholar

I. Shanthi 08.08.23

Assistant Librarian

D. Scm 8/8/23

Research Guide

Date: 08-08 2023

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Amala Jayanthi M
Assignment title:	New 2022
Submission title:	An Integrated Framework for Students' Academic Performan...
File name:	19phcsp001_Modified_Full_Thesis_after_plagiarism_changed...
File size:	2.2M
Page count:	147
Word count:	23,623
Character count:	142,640
Submission date:	08-Aug-2023 10:52AM (UTC+0530)
Submission ID:	2142970919

An Integrated Framework for Students' Academic Performance Prediction Based on Affective and Cognitive State Using Mixed Numeric and Categorical Data

i

An Integrated Framework for Students' Academic Performance Prediction Based on Affective and Cognitive State Using Mixed Numeric and Categorical Data

by Amala Jayanthi M

Submission date: 08-Aug-2023 10:52AM (UTC+0530)
Submission ID: 2142970919
File name: 19phcsp001_Modified_Full_Thesis_after_plagiarism_changed_v4.docx (2.2M)
Word count: 23623
Character count: 142640

An Integrated Framework for Students' Academic Performance Prediction Based on Affective and Cognitive State Using Mixed Numeric and Categorical Data

ORIGINALITY REPORT

6%	3%	5%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ijeecs.iaescore.com Internet Source	1%
2	Amala Jayanthi M, Elizabeth Shanthi I, Lakshmana Kumar Ramasamy. "Recommender System for Predicting Students' Academic performance in association with Cognitive state and Affective state using Sentiment Analysis and Association Rule Mining on the closed ended questionnaire", 2023 9th International Conference on Information Technology Trends (ITT), 2023 Publication	1%
3	downloads.hindawi.com Internet Source	1%
4	Amala Jayanthi M, Elizabeth Shanthi I. "Quest_SA: Preprocessing Method for Closed-Ended Questionnaires Using Sentiment	<1%