

Open Issues, Research Challenges, and Survey on Education Sector in India and Exploring Machine Learning Algorithm to Mitigate These Challenges

¹K Rajesh Kannan, ¹K T Meena Abarna, ²S Vairachilai

¹Department of Computer Science and Engineering, Annamalai University, Chidambaram, Tamil Nadu, India

²School of Computer Science and Engineering (SCSE), VIT Bhopal University, Madhya Pradesh, India.

Email ID: rajeshlpm88@gmail.com

Abstract— The nation's core sector is education. But dealing with problems in educational institutions, particularly in higher education, is a challenging task. The growth of education and technology has led to a number of research challenges that have attracted significant attention as well as a notable increase in the amount of data available in academic databases. Higher education institutions today are worried about outcome-based education and various techniques to assess a student's knowledge level or capacity for learning. In general, there are more contributors in the academic field than there are authors. Research is being done in this field to determine the best algorithm and features that are crucial for predicting the future outcomes. This survey can help educational institutions assess themselves and find any gaps that need to be filled in order to fulfil their purpose and vision. Machine Learning (ML) approaches have been explored to solve the issues as higher education systems have grown in size.

Keywords- Hybrid Machine Learning Algorithms, Cross Validation, Artificial Intelligence, Academic Performance.

I. INTRODUCTION

The pandemic situation has made it difficult for students to be overburdened, which has significantly slowed global education growth. Basically, different types of students are present in these classrooms. While some students are self-motivated, others could struggle to understand the principles. It might be challenging to assess students' performance and learning potential. Although the fact that many of institutions rely their assessment of a student's performance on their grade point average, this method relies on consistent pedagogical approaches and on-going internal evaluation. The most recent technological developments also enable the students to learn more efficiently by utilizing online tools and resources as needed. The majority of educational institutions are using a variety of Information and Communication Technology (ICT) technologies to access and assess student performance as part of an efficient teaching and learning process in this pandemic environment. The Learning Management System (LMS) is one of the tool to provide E-Learning at free of cost. Additionally, LMS enables better management, records user activity, and stores and processes massive volumes of educational content. To extract the important patterns from the databases, the researchers are concentrating on examining the LMS data, which has rapidly increased. In this article, we focused on academic-related topics like dropout factor prediction, academic performance analysis, and a system for recommending elective courses based on real-world features.

The opportunity to lead the country toward success is provided by education. The educational institutions' core principle is to provide their students a quality education in order to develop their careers. To track their students' progress, several educational institutions have adopted cutting-edge technologies like artificial intelligence. Numerous criteria, including midterm exams, assignments, and slip tests, among others, are used to address the numerous research problems in the academic sector. However, this problem is handled by a variety of special features. This paper's objective is to look at the issues that are gaining significant attention in the academic sector.

Many institutions started exploring with online learning platforms during the pandemic. An institution must upgrade its information communication technology resources, strengthen the required technical interventions, and train its faculty and staff to develop shared resources and exchange platforms for conducting discussions, challenges, and solutions before deploying an online learning platform. The government has also launched an open platform, such as the National Program on Technology Enhanced Learning, Massive Open Online Course (MOOC) Platform, and Study Webs of Active Learning for Young Aspiring Minds (SWAYAM), and many others, to support educational institutions and ensure achievement. Aiming for a 50 percent Gross Enrolment Ratio (GER) in higher education by 2035 is the goal of India's NEP-2022 [8] (New Education Policy).

II. RELATED WORKS

The various studies have been reviewed and many different (stand-alone) machine learning algorithms have been implemented to analyse and predict student progress in recent years. Linlin Zhang and Kin Fun Li [1] discuss the eighty eight papers/projects which were published in the field of academics from 2013 to 2017. In which, each and every project is implemented with standalone machine learning algorithms only.

On a set of educational data, three models based on an ensemble of regressors are used. These models are studied using a novel [2] data mining applications based Cross Industry Standard Procedure approach. The criteria taken into account by the higher education census database include inadequate motivation, personal, and socioeconomic issues, discontent with the course or institution, learning issues related to teaching approaches, lack of course knowledge, and degree of prior study. In the modelling step, a sample set of the original data is produced using the bagging approach. In this project academic performance of the kids is not taken into consideration.

The process of filtering the vast amounts of data from the database to find patterns is known as academic analytics. Mathematical evaluation [3] is typically conducted using histograms and kernel estimations of test results for graduates and dropouts. One-hot encoding was employed to convert categorical variables to numerical variables in order to prepare the entire dataset for modelling and imputation. The gradient boosted tree is trained to predict dropout. However, the dataset falls into four major categories. They are:

- Metrics of high school performance
- University Program
- Personally Identifiable/Confidential Information
- First year related information

In this study, a framework for data mining applications based on Cross Industry Standard Procedure and the techniques were applied in order to investigate the trends in the institutional data used in this project [4]. The characteristics can be divided into four categories: (i)

academic performance, (ii) financial factors, (iii) prior educational history, and (iv) demographic aspects. The data was combined into a single flat file after being collected from numerous data sources. The parameters considered in this paper may not be suitable for other branches as they are specifically selected for computing branches. In addition, [5] discussed the analysis of data from various data from the databases could be utilised for the decision making process by the institute. However, most of the studies were based on the standalone machine learning algorithms only.

Martn Sols [6] presented a model where the data must be split into at least two parts: one for training the model and the other for gauging its predictive power. A student is deemed a dropout if they have gone at least two academic years without enrolling and have not yet graduated, as opposed to a non-dropout if they are currently enrolled or have completed their studies between 2011 and 2016. Since it is unknown in advance whether they will graduate or drop out, the goal in the second perspective is to remove the noise of active students when training the algorithm. The third perspective seeks to remove noise from the dropout's earlier semesters on the grounds that the most current semester contains the most recent data to determine whether or not the dropout will occur. The fourth perspective seeks to reduce noise from both current students and dropouts from prior semesters. This project works with several models which makes it time taking and falls in the category of trial and error. Data taken in this processor is not completely noiseless. So, this may result in more processing of data.

Ihsan A. Abu Amra [7] proposed a model based on the classification algorithms on the academic dataset taken from the education ministry database Gaza. This Research work can enable the teachers to do the evaluation to improve each and every student. Despite only taking into account a small number of criteria for predicting student performance; the Naive Bayes classifier needs a vast number of data to produce reliable results. Table I shows the detailed study about Academic success prediction for last seven years. The All India Higher Education Survey (AISHE) was established by the Ministry of Human Resources Development

Table 1. Review of students' academic performance

| Ref Paper | Author | Year | Problem Statement | Algorithms Used |
|-----------|----------------------|------|---|-----------------|
| 9 | Praveena Chakrapani | 2022 | Student Performance Prediction | DT, RF ,SVM |
| 10 | Dindar Mikaeel Ahmed | 2021 | Student Performance Prediction | LR, SVM |
| 11 | Jitendra Darji | 2021 | Student Performance Prediction based on their Learning Experience | DT, KNN, SVM |
| 12 | Aaditya Bhusal | 2019 | Student Performance Prediction | DT, RF ,SVM, NN |
| 13 | Pushpa S K | 2017 | Student Performance Prediction | RF, SVM, NB |
| 14 | Elaf Abu Amrieh | 2016 | Student Performance Prediction | DT, NB, ANN |

| | | | Using Ensemble Methods | |
|----|-----------------------|------|---|--------------------------|
| 15 | Durgesh Ugale1 | 2020 | Student Performance Prediction with Data Mining | DT, SVM, KNN, NB |
| 16 | Mudasir Ashraf | 2020 | An Intelligent Prediction Model Using Data Mining | DT, RF, NB, KNN |
| 17 | Vaibhav Kumar | 2019 | Student Result Prediction | GLM, MLP,GBM, RFM, DNN |
| 18 | Fergie Joanda Kaunang | 2018 | Student Performance Prediction using Data Mining | DT, RF |
| 19 | Leila Ismail | 2021 | Student Performance Prediction Using ML Models | DT, NB, ANN, RF, SVM |
| 20 | Eyman Alyahyan | 2020 | Academic Success Prediction in Higher Education | DT, NB, ANN, KNN |
| 21 | Kongara Deepika | 2018 | Student Performance Prediction | DT, SVM, KNN, NB |
| 22 | K. Sripath Roy | 2018 | Student Career Prediction Using Advanced ML | DT, SVM, XGBoost |
| 23 | Ahmed Mueen | 2016 | Student Performance Prediction using Data Mining | DT, NB, MLP |
| 24 | Sadiq Hussain1 | 2019 | Classification in Educational Dataset Using Data Mining | DT, NN, NB |
| 25 | Raza Hasan | 2020 | Student Performance Prediction | DT, SVM, LR, KNN, NB, NN |
| 26 | Bindhia K. Francis | 2019 | Student Performance Prediction using Hybrid Data Mining | SVM, NB, DT, NN |

Conducted an annual web-based survey to show the status of higher education in the country. We have reviewed [AISHE 2019-2020] thoroughly and it was observed the Gross Enrolment Ratio of the country improves alongside the significant challenges remains especially after the pandemic situation. Figure 1 shows the growth of GER in last five years. The report of National Statistical Office under the Ministry of Statistics and Programme Implementation reveals one in eight students enrolled in either school or higher education quit their education before completing the course[27].

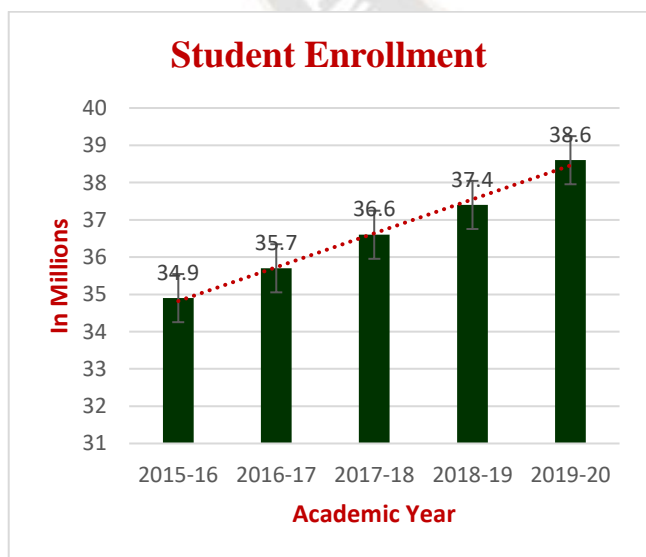


Figure 1. Growth of GER in India.

III. Research Challenges

This section deals with research challenges, which needs to be rectified to retain the current student enrolment in various courses of school/higher education. There are various reasons behind the higher dropout rate such as inadequate school education, family economic conditions, and inadequate access to the education especially in rural areas, and unusual pandemic situations etc.

A. Academic Performance Prediction

Academic achievement is one of the most significant aspect considered by all educational institutions when appraising the students. In addition, Institutions are increasingly focusing on providing students with quality education in a competitive world [10]. Educational institutions place a strong emphasis on both extracurricular activities and academic achievement. They are keep on monitoring of how well students are doing in various area in the sector and in which areas they require lot of focus.

The ability to analyse and student performance prediction is an integral part of the education sector. Extracting the useful patterns from the educational database can be done using several Machine Learning Algorithms [1], such as Decision tree (DT), Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), Naive Bayes Classifiers (NB), K-Nearest Neighbors (KNN), Neural Networks (NN), and

Gradient Boost Model (GBM)[17]. However, all these algorithms belong to either classification or clustering. There are different algorithms for predicting academic success of students, but a hybrid machine learning approach that combines both clustering and classification algorithms are more accurate than the existing algorithms.

B. Employment Status/Career Prediction in Higher Education

In many university students aspire to have lucrative careers. A high-profile job or internship during their stay on campus will be advantageous. [28] The placement activities that are carried out at the right time can be very beneficial to undergraduates from any major. An important indicator of student employment Prospects can help the university stabilize it better Output from college students that meet the true needs of society.

Currently, most of the universities have internship programs. It is considered an integral part of the curriculum. Internships are defined as "short-term, on-the-job experience during which students receive instruction and practise in certain fields of interest or careers". Internship program facilitates transition from college to real world workplace because they are an effective way of understanding what they have learned as part of curriculum and how they are going to meet the industry requirements. Career is a sequence of developments or advances everyone experienced it. This career's [29] work title and something associated with the professional world first surfaced. The advancement of the life journey as it is experienced through individuals is the most crucial component of a career.

C. Student Dropout Prediction

Dropout rates among pupil are a challenging and important problem in the education system, having detrimental

D. Elective Track /Course Recommender System in Higher Education

Universities frequently need in recommending that students are enrolled in particular courses. If students may choose to study what they enjoy or are interested in, it not only aids in their decision-making but also enhances their academic performance. Recommender systems are emerging in academic institutions [31], and they have a big influence on students' study plans and grades. Today learners' primary aspect is that they require customized access to the Industry Standard Elective Courses based on their needs and preferences. The recommender system is used to analyse the data automatically in accordance with user priority and the best choice is suggested to succeed their academic

effects on learners, academic institutions, financial resources, and society as a whole. This study looked at the overall and comparative effects of student (personal), academic, and socioeconomic factors on dropout. Fig 2 shows Dropout Rate by level of School education and gender, 2020-21 [30].

Due to school closures, irregular online lessons, and the loss of revenue for low-income families, the dropout rate may have significantly increased in recent pandemic time. High rates of school abandonment have a negative impact on people's ability to pursue further education and careers. The success of a nation rests, among other things, on its capacity to generate graduates from higher education who can advance the nation. The dropout problem is one that more and more institutions are trying to address with artificial intelligence's [3] potential to predict dropout as early as possible.

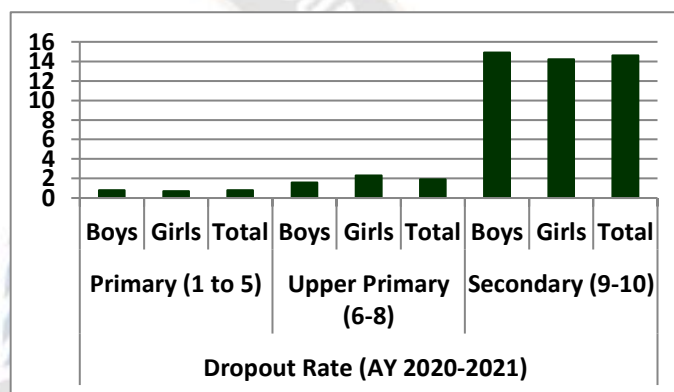


Figure 2. Dropout rate by level of school education and gender

Depending on the data used, Machine Learning can also be employed to develop predictive models that predict student dropout. Some of them use the students' own, their families', and their environment's personal information; others examine students' behavior and interaction data which is stored by online educational platforms and tools.

career. Here the Information was prioritized, linked to other items, and users were given helpful suggestions based on their interests by the recommender system [32]. The most often used techniques for recommendation include rule mining techniques, collaborative filtering, and content-based filtering

E. Graduate Admission Prediction in Higher Education

Every year large number of Indian Students applying for graduate admissions to the various universities across the globe [34]. In Which the United States of America (USA), Canada, and the United Kingdom (UK) were the top three countries for education in 2021, according to statistics obtained from the Bureau of Immigration (BoI). Prospective graduate students are frequently struggle with the decision of which universities to attend while applying to graduate

programmes. While there are many predictors and consultancies available to help a student, they are not always accurate because admissions decisions are based on historical admissions data and crucial qualitative elements like the Letter of Recommendation [33] (LOR) and the Statement of Purpose (SOP). In this direction, a machine learning model can be created to accurately predict a student's chance of admission to the institution of his or her choice. Figure 3 shows the number of Indian students who left India for pursuing higher education in the current year and the two years before according to data from the Bureau of Immigration.

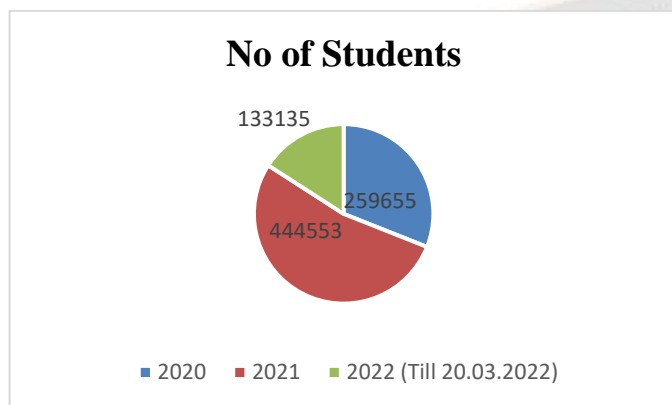


Figure 3. Indian students pursuing higher education in across the world

IV. CONCLUSION

This survey has shown a growing interest of the multiple research domains in academic sector. The aim is to analyse these research challenges to gain a better understanding and identify the appropriate Solutions. Hybrid machine learning algorithms with cross validation is recommended to address all the research challenges discussed so far. This will enable the institutions to produce quality graduates and modify their curriculum to enhance their teaching and learning process to build the better nation in future. We believe that this survey will give readers a chance to re-evaluate with state-of-the-art techniques.

REFERENCES

[1] Zhang, L. & Li, K.F. (2018) Education analytics: Challenges and approaches 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp. 193–198 [DOI: 10.1109/WAINA.2018.00086].

[2] da Silva, P.M., Lima, M.N.C.A., Soares, W.L., Silva, I.R.R., Fagundes, R.A.de A. & de Souza, F.F. (2019) Ensemble regression models applied to dropout in higher education 8th Brazilian Conference on Intelligent Systems (BRACIS), pp. 120–125 [DOI: 10.1109/BRACIS.2019.00030].

[3] Kiss, B., Nagy, M., Molontay, R. & Csabay, B. (2019) Predicting dropout using high school and first-semester academic achievement measures 17th International Conference on Emerging eLearning Technologies and Applications (ICETA), pp. 383–389 [DOI: 10.1109/ICETA48886.2019.9040158].

[4] Naseem, M., Chaudhary, K., Sharma, B. & Lal, A.G. (2019) Using ensemble decision tree model to predict student dropout in computing science IEEE Asia and the Pacific Conference on Computer Science and Data Engineering (CSDE), pp. 1–8 [DOI: 10.1109/CSDE48274.2019.9162389].

[5] Campbell, J.P., DeBlois, P.B. & Oblinger, D.G. (2007) Academic analytics: A new tool for a New Era. *Educause Review*, 42, 40–57.

[6] Solis, M., Moreira, T., Gonzalez, R., Fernandez, T. & Hernandez, M. (2018) “Perspectives to Predict Dropout in University Students with Machine Learning,” IEEE international work conference on bioinspired intelligence (IWOB), pp. 1–6 [DOI: 10.1109/IWOB.2018.8464191].

[7] Abu Amra, I.A. & Maghari, A.Y.A. (2017) Students performance prediction using KNN and Naïve Bayesian 8th International Conference on Information Technology (ICIT), pp. 909–913 [DOI: 10.1109/ICITECH.2017.8079967].

[8] www.education.gov.in/sites/upload_files/mhrd/files/NEP_Final_English_0.pdf.

[9] Chakrapani, P. & C.D. (2022) Academic performance prediction using machine learning: A comprehensive & systematic review International Conference on Electronic Systems and Intelligent Computing (ICESIC), pp. 335–340 [DOI: 10.1109/ICESIC53714.2022.9783512].

[10] Ahmed, D.M., Abdulazeez, A.M., Zeebaree, D.Q. & Ahmed, F.Y.H. (2021) Predicting University’s students’ performance based on machine learning techniques IEEE International Conference on Automatic Control. & Intelligent Systems (I2CACIS), pp. 276–281 [DOI: 10.1109/I2CACIS52118.2021.9495862].

[11] Darji, J. & Nakrani, T. (2021) Enhance student learning experience by using machine learning to predict student performance in advance. *International Research Journal of Modernization in Eng. Tech. & Sci*, 03.

[12] Bhusal, A. (2021). Predicting Student’s Performance Through Data Mining. [DOI: <https://doi.org/10.48550/arXiv.2112.01247>]

[13] Pushpa, S.K., Manjunath, T.N., Mrunal, T.V., Singh, A. & Suhas, C. (2017) Class result prediction using machine learning International Conference On Smart Technologies For Smart Nation (SmartTechCon), pp. 1208–1212 [DOI: 10.1109/SmartTechCon.2017.8358559].

[14] Amrieh, E.A., Hamtini, T. & Aljarah, I. (2016) Mining Educational Data to Predict Student’s academic Performance using Ensemble Methods. *International Journal of Database Theory and Application*, 9, 119–136 [DOI: 10.14257/ijtda.2016.9.8.13].

[15] Ugale, D., Pawar, J., Yadav, S. & Raut, D.C. (2020) Student performance prediction using data mining

- techniques. *International Research Journal of Engineering and Technology*, 07.
- [16] Ashraf, M., Zaman, M. & Ahmed, M. (2020) An intelligent prediction system for educational data mining based on ensemble and filtering approaches. *Procedia Computer Science*, 167, 1471–1483 [DOI: 10.1016/j.procs.2020.03.358].
- [17] Kumar, V. & Garg, M.L. (2019) Comparison of machine learning models in student result prediction. *Advances in Intelligent Systems and Computing International Conference on Advanced Computing Networking and Informatics*, 870, 439–452 [DOI: 10.1007/978-981-13-2673-8_46].
- [18] Jayaprakash, P. O. ., Raja, K. T. ., Dharinya, S. S. ., & Janaki, N. . (2023). Artificial Bee Colony Optimized Recurrent Neural Network-Based Port Container Throughput Forecast. *International Journal of Intelligent Systems and Applications in Engineering*, 11(1), 12–17. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/2438>
- [19] Kaunang, F.J. & Rotikan, R. (2018) Students' academic performance prediction using data mining Third International Conference on Informatics and Computing (ICIC), Vol. 2018, pp. 1–5 [DOI: 10.1109/IAC.2018.8780547].
- [20] Ismail, L., Materwala, H. & Hennebelle, A. (2021) Comparative Analysis of Machine Learning Models for Students' Performance Prediction. *Advances in Intelligent Systems and Computing, International Conference on Advances in Digital Science*, 1352, 149–160 [DOI: 10.1007/978-3-030-71782-7_14].
- [21] Alyahyan, E. & Düşteğör, D. (2020) Predicting academic success in higher education: Literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17, 3 [DOI: 10.1186/s41239-020-0177-7].
- [22] Kongara, D. & Sathyanarayana, N. (2019) Relief-F and budget tree random forest based feature selection for student academic performance prediction. *International Journal of Intelligent Engineering and Systems*, 12. 30-39, 10.22266/ijies2019.0228.04.
- [23] Roy, K. & Roopkanth, K. & Teja, V. & Bhavana. & Priyanka, J.. (2018). Student Career Prediction Using Advanced Machine Learning Techniques. *International Journal of Engineering and Technology(UAE)*. 7. 26-29. 10.14419/ijet.v7i2.20.11738.
- [24] Mueen, A., Zafar, B. & Manzoor, U. (2016) Modeling and predicting students' academic performance using data mining techniques. *International Journal of Modern Education and Computer Science*, 8, 36–42 [DOI: 10.5815/ijmecs.2016.11.05].
- [25] Hussain, Sadiq & Atallah, Rasha & kamsin, Amirrudin, Hazarika & Jiten (2019). Classification, Clustering and Association Rule Mining in Educational Datasets Using Data Mining Tools: A Case Study. 10.1007/978-3-319-91192-2_21.
- [26] Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K.U. & Sattar, M.U. (2020) Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques. *Applied Sciences*, 10, 3894 [DOI: 10.3390/app10113894].
- [27] Francis, B.K. & Babu, S.S. (2019) Predicting academic performance of students using a hybrid data mining approach. *Journal of Medical Systems*, 43, 162 [DOI: 10.1007/s10916-019-1295-4] [PubMed: 31037484].
- [28] <https://theeducationdaily.com/2021/07/dropouts-higher-education-india>.
- [29] Saidani, O., Menzli, L.J., Ksibi, A., Alturki, N. & Alluhaidan, A.S. (2022) Predicting student employability through the internship context using gradient boosting models. In: *IEEE Access*, 0, 46472–46489 [DOI: 10.1109/ACCESS.2022.3170421].
- [30] Rattan, V., Sharma, S., Mittal, R. & Malik, V. (2021) Applying SMOTE with decision tree classifier for campus placement prediction International Conference on Computing, Communication and Green Engineering (CCGE), Vol. 2021, pp. 1–6 [DOI: 10.1109/CCGE50943.2021.9776360].
- [31] https://udiseplus.gov.in/assets/img/dcf2021/UDISE+2020_21_Booklet_English.pdf.
- [32] Alamir, M.B. & S. (2019) Larabi-Marie-Sainte and D. Alsaleh. A Student-Centric Prototype for Courses Recommendation in Universities' Students Portal 12th International Conference on Developments in eSystems Engineering (DeSE), Vol. 2019, pp. 591–596 [DOI: 10.1109/DeSE.2019.00112].
- [33] Bhumichitr, K., Channarukul, S., Saejiem, N., Jiamthapthaksin, R. & Nongpong, K. (2017) Recommender Systems for university elective course recommendation 14th International Joint Conference on Computer Science and Software Engineering (JCSSE), Vol. 2017, pp. 1–5 [DOI: 10.1109/JCSSE.2017.8025933].
- [34] Khan, M.A., Dixit, M. & Dixit, A. (2020) Demystifying and anticipating Graduate School admissions using machine learning algorithms 9th International Conference on Communication Systems and Network Technologies (CSNT), Vol. 2020. IEEE Publications, pp. 19–25 [DOI: 10.1109/CSNT48778.2020.9115788].
- [35] <https://www.mea.gov.in/rajya-sabha.htm?dtl/35106/>.