ISSN: 2633-4828 Vol. 5 No. S4 (July - Aug 2023)

International Journal of Applied Engineering & Technology

REVAMPING CURRICULUM USING DATA MINING AND PYTHON: SYSTEMATIC REVIEW

[1]Ritika Awasthi and [2]Dr. Arvind Tiwari
[1]Research Scholar, Dr. C. V. Raman University
[2]Associate Professor, Dr. C.V. Raman University
[1]ritikaawasthi1612@gmail.com and [2]arvindtiwari@cvru.ac.in

ABSTRACT
After the National Education Policy's (NEP) commencement, colleges and universities worldwide are required to review their curriculum to ensure that they meet modern needs and satisfy the expectations of students and all other stakeholders. This systematic review's subject incorporates Python programming and data mining tools in curriculum reform techniques post-NEP-2020. The study performs a thorough literature analysis and looks at the approaches, materials, and findings associated with employing Python and data mining to reform the curriculum after the NEP. The review illustrates the broad spectrum of possibilities for data mining, from educational materials optimization to student assessment of performance. It also describes how Python plays a crucial role in enabling modifying data, visualization, and predictive modeling in educational settings. Relevant findings reflect how well these technologies work to improve student engagement, curriculum importance, and adaptability. Additionally, the evaluation reveals novel opportunities, challenges, and potential for more research and implementation.

Keywords: Data Mining, Python Programming, Support Vector Machine, Curriculum redesigning, NEP.

1. INTRODUCTION
Higher education in India is considered to be the third largest education system in all over world, next to US and China. University Grant Commission is the main governing body at university or college level which imposes its standard rules and suggestions and help in coordination between Centre and state higher education authority. Indian education system is very comprehensive and revolutionary with respect to approachability and reforms in rule and regulations, putting values and rate of movement to a particular course. Following are the vision of NEP 2020 for the higher education system:

- Quality higher education must aim to develop good, thoughtful, well-rounded, and creative individuals.
- It must enable an individual to study one or more specialized areas of interest at a deep level, and also develop character, ethical and Constitutional values, intellectual curiosity, scientific temper, creativity, spirit of service, and 21st-century capabilities across a range of disciplines including sciences, social sciences, arts, humanities, languages, as well as professional, technical, and vocational subjects.
- A quality higher education must enable personal accomplishment and enlightenment, constructive public engagement, and productive contribution to society.
- It must prepare students for more meaningful and satisfying lives and work roles and enable economic independence.

Curriculum, Assessment and Evaluation are the major tools by which Program Outcomes are attained(Cervantes et al., 2020). APTOBE (A Python tool for Outcome based education) considers all of these together. An Outcome based assessment system comprises of several components. APTOBE was developed to assist faculty in designing examination questions with respect to Bloom's Taxonomy cognitive levels. APTOBE encourages faculty to adhere to the Taxonomy by displaying the keywords of different levels, along with the syllabus, while designing the question paper.(Reddy Tetali et al., 2016) By using this tool, with SVM this study will revamp a curriculum which will be beneficial for each and every stakeholder of the higher education structure. SVM is one of the most used Data Mining Techniques. The SVM provides up to 98% accurate result with python. It is a supervised

learning method and can be used for both regression and classification. SVM Classifier is known for maximum accuracy and Minimum Root Mean Square error (RMSE).

2. APPEARANCE AND JUSTIFICATION OF PROBLEM

India's National Education Policy 2020 (NEP 2020) has been the focus of numerous critical investigations. The policy's emphasis on a comprehensive and multidisciplinary approach, with an emphasis on holistic growth and across the globe have significance has been emphasized by these studies (2022). As we know one the feature of NEP (2020) is that to revamp the curriculum, pedagogy and assessment of students and this is the main focus of this study. Moreover, A curriculum should be evaluated time to time and redesign to reflect the changing technology in curriculum. Information and experience should be added, deleted and adjusted to reflect the prevailing beliefs or findings. Students change because of impact of social media and many other technologies modification. As educator we should also update our knowledge and skills. Moreover, as stated above curriculum should also meet NEP objectives, not only academically but professionally as well. As technology evolved drastically in past few years so the curriculum should also be up to the mark to achieve all the goals and objectives.

Since now there is new NEP (2020) which brings major changes in course outline, there is need to be updated our knowledge and skills to achieve the motive of NEP. The main motto of is Educate, Encourage and Enlightened. To achieve this concept there is requirement of redesigning of curriculum so that it can be more fruitful for students and other stakeholders of higher education system. But there are also a lot of obstacles the strategy must overcome, such as the necessity to deal with the education sector's problems and guarantee equitable access to high-quality education (2023).

3. GOALS OF THE RESEARCH

- To develop SVM using Python for redesign curriculum to fulfill the knowledge and need of learners, educators as well as industries.
- To develop a Python-based module for providing a suitable curriculum with the help of SVM for different groups of students and for different pedagogies.
- A Python-based software education paradigm emphasizing cognitive thinking and problem-solving for non-computer majors.
- To develop a Program specific outcome (PSO) oriented curriculum.
- To develop a Program Educational Objective (PEO) based curriculum. The development of a computer science education curriculum based on Program Academic Goals (PEOs) has been the subject of many research projects.

4. METHODOLOGY OF THE STUDY

Following are the steps of methodology which usually followed:

4.1. Data collection: To achieve the objective, first of all data collected from students of higher education from different colleges and university. Data repository can also be one source for data collection. Size of sample can be varied in between 500 to 1000 including survey of students, faculties and administrative staff.

4.2. Data authenticated by taking consent of stakeholders of colleges and universities and data collection will be done through google sheet in form of questionnaire in which date will be mentioned.

4.3. After collection of data, next step is data cleaning. Following are the steps of data cleaning:

- Remove duplicate data: when data are collected from different places and multiple sources there are chances of duplicate data collection. Remove all duplicate and irrelevant data from the collection to make the data set more manageable and accurate.

- Remove structural error: Error in context, syntax, and terminology removed.
- Remove outliers: There may be some outliers that exceed the measurement criteria of this study. For example, there may be data that will not match the attributes of the study; those are removed in this step.
- Handling missing data: In this step requirement of missing data is fulfilled by again collecting data for missing information or some information can be altered.

4.4 This step is followed by data processing. By this process, raw data is converted in to meaningful information in the form of a report and chart and graph so that it would be easy to fetch result from gathered data.

4.5 With a focus on enhancing teaching quality as well as educational efficiency, data processing in education using data mining and Python is a rapidly growing field(Wang et al., 2022a). This strategy makes use of techniques for data mining such as social area networking, forecasting, categorization, relationship mining, and clustering(Sachin et al, 2012). These methods are used in education in a range of ways, such as analyzing student trends and behavior(Aher s , 2011). Because of its versatility and simplicity, Python is very helpful in such a scenario(Wang et al., 2022b).

4.6 Next step is data transformation with the help of different prediction algorithm and Education Data Mining models. This process is done by using data and by applying different attributes to algorithms.

4.7 Final step will be producing result and try to meet objectives of the study along with the analysis of proposed model.

5. TOOLS AND TECHNIQUES:

There has been growing interest in the integration of data mining towards Python curriculum redesign. The potential of using data mining for educational situations has been recognized by(Agnihotri et al., 2016a) and (Chu et al., 2018), with Agnihotri focusing on the use of technologies such as big data and Apache Spark for analysis. Both (Baker Ed. et al., 2010)and (Nguyen & Ha, 2019) offer practical instances of data mining techniques applied in educational situations.

- **Big Data Analysis:** Big data is an approach of handling huge quantities of unstructured and structured data that are difficult to manage in traditional databases by retrieving, gathering, managing, and then evaluating it (Belcastro et al., 2019).To assist developers and customers choose the best solution, the method analyzes the features of standard programming models for Big Data analysis with primary systems and evaluates them. The contrast is based on four classification criteria.
- **Clustering Algorithm:** The artificial intelligence (AI) mathematics teaching and learning materials were selected because both of them demonstrate how AI utilizes mathematics to solve issues and are suitable for the student's level of knowledge. The process comprised establishing standards for the creation of teaching-learning materials, producing four lessons of materials for instruction in mathematics for artificial intelligence, and revising the materials under responses from teachers and testing in the field(Zheng, 2012).
- **Association Rule:** The results when using these strategies can significantly impact program goals and provide direction. To forecast student performance and identify connections between educational variables, data mining techniques are crucial. The rules' level of interest was evaluated using specific metrics, and examples of significant course conjunctions were discovered(Alangari et al., 2020).

6. CONCLUSION

In order to sum up, the execution of methods for data mining has great potential for reusing curricula under the National Education Policy (NEP). Data mining may significantly enhance learning results due to its ability to derive insightful inferences and patterns from enormous data sets. Support Vector Machines, or SVM, and Python's many data analysis bundles provide powerful tools for identifying patterns and predictive modeling in educational environments. In addition, such as data mining techniques within STEM (Science, Technology,

Engineering, and Mathematics) curriculum encourages a more quantitative and information-driven style of teaching. This paper uses a comprehensive data mining approach for curriculum revamping, that is made feasible by the tools used, which involve the analysis of big data devices, Apache Spark for extensible process of data, algorithms for clustering for grouped similar information about education, and association rule mining techniques. Educational institutions might increase the overall effectiveness of instruction in line with the NEP's ambitions, improve teaching practices, and respond to the evolving needs of students through the use of these methods and instruments.

7. FUTURE SCOPE:

Under the encouragement of strong ethical data exploitation applications, potential applications of mining data and Python post-NEP curriculum reform might ultimately include dynamically adaptive learning systems tailored to the needs of particular students. These technologies, which emphasize individual privacy and confidentiality while providing scalable personalized learning experiences, can completely transform conventional approaches to learning because they are powered by immediate information analysis. Additional research ought to concentrate on performing longitudinal impact studies to evaluate the long-term effects of data-driven solutions on learning outcomes and readiness for employment. The results of these investigations would help foster the creation of evidence-based instructional policy and practice, ensuring ongoing enhancement of data-driven curriculum development techniques.

REFERENCES

- Agnihotri, L., Mojarad, S., Lewkow, N., & Essa, A. (2016b). Educational Data mining with python and apache spark: A hands-on tutorial. ACM International Conference Proceeding Series, 25-29-April-2016. https://doi.org/10.1145/2883851.2883857.
- Alangari, N., & Alturki, R. (2020). Association rule mining in higher education: A case study of computer science students. In EAI/Springer Innovations in Communication and Computing. https://doi.org/10.1007/978-3-030-13705-2_13.
- Al-khresheh, M. H. (2024). Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching. Computers and Education: Artificial Intelligence, 6. https://doi.org/10.1016/j.caeai.2024.100218
- Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. D. (2020). An Introduction to Machine Learning. Clinical Pharmacology and Therapeutics, 107(4). https://doi.org/10.1002/cpt.1796
- Baker Ed., R. S. J. d., Merceron Ed., A., & Pavlik Jr., Ed., P. I. (2010). [Proceedings of the] International Conference on Educational Data Mining (EDM) (3rd, Pittsburgh, PA, July 11-13, 2010). In International Working Group on Educational Data Mining.
- Bansal, S., Bansal, A., & Dalrymple, O. (2015). Outcome-based education model for computer science education. Journal of Engineering Education Transformations, 28(2).
- Bart, A. C., Friend, M., Sarver, A., & Cox, L. (2019). Pythonsneks: An open-source, instructionally-designed introductory curriculum with action-design research. SIGCSE 2019 - Proceedings of the 50th ACM Technical Symposium on Computer Science Education, 307–313. https://doi.org/10.1145/3287324.3287428
- Belcastro, L., Marozzo, F., & Talia, D. (2019). Programming models and systems for Big Data analysis. International Journal of Parallel, Emergent and Distributed Systems, 34(6). https://doi.org/10.1080/17445760.2017.1422501
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408. https://doi.org/10.1016/j.neucom.2019.10.118

- Chen, P., Chen, I., & Verma, R. (2011). Designing an undergraduate data mining course by matching teaching strategies with student learning styles. Journal of Computing Sciences in Colleges, 26(4).
- Chopra, D., & Khurana, R. (2023). Introduction to Machine Learning with Python. In Introduction to Machine Learning with Python. https://doi.org/10.2174/9789815124422123010l
- Douglas, J. V., Bianco, S., Edlund, S., Engelhardt, T., Filter, M., Günther, T., Hu, K. M., Nixon, E. J., Sevilla, N. L., Swaid, A., & Kaufman, J. H. (2019). STEM: An open-source tool for disease modeling. Health Security, 17(4). https://doi.org/10.1089/hs.2019.0018
- Durugkar, S. R., Raja, R., Nagwanshi, K. K., & Kumar, S. (2022). Introduction to data mining. In Data Mining and Machine Learning Applications. https://doi.org/10.1002/9781119792529.ch1
- Guenther, N., & Schonlau, M. (2016). Support vector machines. In The Stata Journal (Vol. 16, Issue 4).
- Guleria, P., & Sood, M. (2014). Data Mining in Education: A Review on the Knowledge Discovery Perspective. International Journal of Data Mining & Knowledge Management Process, 4(5), 47–60. https://doi.org/10.5121/ijdkp.2014.4504
- Gupta, B., Gupta, B. L., & Choubey, A. K. (2021). Higher Education Institutions-Some Guidelines for Obtaining and Sustaining Autonomy in the Context of Nep 2020. In International Journal of All Research Education and Scientific Methods (IJARESM) (Vol. 9, Issue 1).
- Huang, Y., Zhao, J., Qiang, Y., Hou, T., Ren, X., & Sun, H. (2021). The reform and exploration of intelligent PYTHON language teaching. ICCSE 2021 - IEEE 16th International Conference on Computer Science and Education. https://doi.org/10.1109/ICCSE51940.2021.9569670
- Kathi*, S., Naidu*, A. R., & Rangaiah****, B. (n.d.). A Critical Analysis Of India's National Education Policy (2020): Insights Into Educational Transformation. In Journal of Positive School Psychology (Vol. 2022, Issue 9). http://journalppw.com
- Kaur, P., Singh, M., & Josan, G. S. (2015). Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. Procedia Computer Science, 57. https://doi.org/10.1016/j.procs.2015.07.372
- Liu, X., Raj, R., Reichlmayr, T., Liu, C., & Pantaleev, A. (2013). Incorporating service-oriented programming techniques into undergraduate CS and SE curricula. Proceedings - Frontiers in Education Conference, FIE. https://doi.org/10.1109/FIE.2013.6685056
- Magnisalis, I., Demetriadis, S., & Karakostas, A. (2011). Adaptive and intelligent systems for collaborative learning support: A review of the field. In IEEE Transactions on Learning Technologies (Vol. 4, Issue 1, pp. 5–20). https://doi.org/10.1109/TLT.2011.2
- Mahanama, B., Mendis, W., Jayasooriya, A., Malaka, V., Thayasivam, U., & Umashanger, T. (2018). Educational data mining: A review on data collection process. 18th International Conference on Advances in ICT for Emerging Regions, ICTer 2018 - Proceedings. https://doi.org/10.1109/ICTER.8615532
- Mishra, A., Bansal, R., & Singh, S. N. (2017). Educational data mining and learning analysis. Proceedings of the 7th International Conference Confluence 2017 on Cloud Computing, Data Science and Engineering, 491–494. https://doi.org/10.1109/CONFLUENCE.2017.7943201
- Nguyen, T. N. D., & Ha, V. T. T. (2019). AN OVERVIEW OF EDUCATIONAL DATA MINING. Scientific Journal of Tra Vinh University, 1(1). https://doi.org/10.35382/18594816.1.1.2019.88

- Ouraiba, E. A., Chikh, A., Taleb-Ahmad, A., & El Yebdri, Z. (2009). Automatic personalization of learning scenarios using SVM. Proceedings - 2009 9th IEEE International Conference on Advanced Learning Technologies, ICALT 2009. https://doi.org/10.1109/ICALT.2009.72
- Parack, S., Zahid, Z., & Merchant, F. (2012). Application of data mining in educational databases for predicting academic trends and patterns. Proceedings - 2012 IEEE International Conference on Technology Enhanced Education, ICTEE 2012. https://doi.org/10.1109/ICTEE.2012.6208617
- Reddy Tetali, D., Karthik, A., Tech, R. B., & Rajashekar, C. (2016). A Python Tool for Outcome Based Education (APTOBE). International Journal of Emerging Technologies in Engineering Research (IJETER), 4. www.ijeter.everscience.org
- Research on the teaching reform of Python programming curriculum based on the OBE-CDIO concept. (2022). Frontiers in Educational Research, 5(11). https://doi.org/10.25236/fer.2022.051112
- Roslan, M. H. bin, & Chen, C. J. (2022). Educational Data Mining for Student Performance Prediction: A Systematic Literature Review (2015-2021). International Journal of Emerging Technologies in Learning, 17(5). https://doi.org/10.3991/ijet.v17i05.27685
- Sachin, R. B., & Vijay, M. S. (2012). A survey and future vision of data mining in educational field. Proceedings - 2012 2nd International Conference on Advanced Computing and Communication Technologies, ACCT 2012. https://doi.org/10.1109/ACCT.2012.14
- Takada, K., Miyazawa, Y., Yamamoto, Y., Imada, Y., Tsuruta, S., & Knauf, R. (2013). Curriculum optimization by correlation analysis and its validation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7947 LNCS. https://doi.org/10.1007/978-3-642-39146-0_28
- Tiwari, A. K. (2017). Introduction to machine learning. In Ubiquitous Machine Learning and Its Applications. https://doi.org/10.4018/978-1-5225-2545-5.ch001
- Walia, P. (2020). Paradigm Shift in Pedagogical Practices in Mathematics Classroom: NEP 2020. International Journal of Creative Research Thoughts, 8(12).
- Wang, Y., Xu, L., Wang, Q., Lv, H., & Zhang, Y. (2022a). Educational Data Mining and Learning Analysis System Based on Python. Proceedings - 2022 12th International Conference on Information Technology in Medicine and Education, ITME 2022. https://doi.org/10.1109/ITME56794.2022.00122
- Zhang, M., & Pang, L. (2019). Review of Domestic Application Research of Big Data Mining Technology-SVM in Credit Risk Evaluation. https://doi.org/10.2991/seiem-18.2019.64