International Journal of Engineering Applied Science and Management
ISSN (Online): 2582-6948
Vol. 3 Issue 12, December 2022

A Predictive Analysis of Student Performance System Using Sparse Matrix-Matrix Multiplication Algorithm

Mahalya Chopra[1], Mr. Gajendra Singh[2]
M. Tech Scholar, Computer Science and Engineering, SSSUTMS, Sehore, (MP), India[1]
Assistant Professor, Computer Science and Engineering, SSSUTMS, Sehore, (M.P.), India[2]
Email: mahalyachopra@gmail.com[1]

Abstract: *For higher education, it is crucial for the institution to forecast early student academic performance in order to give students appropriate facilities for improving their academic records and make it simple for the instructor to detect them. Numerous machine learning approaches, including k-nearest neighbour, support vector machines, Naive Bayes classifiers, logistic regression, etc., have been created to enhance student academic performance. In this paper, we propose a method for predicting student performance using a machine learning neural network. The MATLAB toolbox is used to develop this model, and comparison analysis is carried out utilising a variety of measurement factors, including precision, recall, f-measure, and execution time.*

Keywords: *Higher learning, Academic performance, Machine learning, Naïve Bayes classifier, Neural network, MATLAB.*

1. Introduction

Student academic achievement is always a matter of great concern to education stakeholders, especially in today's fast-paced, web-enabled classrooms. High quality teaching stuff, well-designed curriculum, student-centered learning and academic support are heavily impact on student success and help to equalize education background differences [1]. Although academic success is an outcome influenced by many factors, research has shown that students‟ engagement with educationally purposeful activities lowers failure rates in introductory college courses and increases retention. The time and effort that students devote to activities linked to desired learning outcomes, such as active and collaborative learning, communication with academic staff and peers, and involvement in enriching educational experiences, has great impact on academic performance [2]. As virtual learning platforms have become the primary means for delivering learning materials for both online and traditional modes of education, instructors have the chance to program online experiences and monitor students‟ activities. The large amount of data that is stored in the log files of these systems can be used to trace student online activities, such as reading files, taking tests, collaborating with peers, communicating with staff etc. [3]. The need to improve the quality of learning has led many institutions to embrace innovative practices like blended learning and use data mining techniques to identify patterns of behaviors and design interventions. Educational data mining provides instructors the tools to filter out information and create models of students‟ performance that predict success [4]. Considering education, higher education institutions have started to build blended learning approaches into their traditional teaching mechanisms to enhance learning and teaching in terms of instructor-generated as well as learner-generated content [5]. Additionally, popularity of massive online open courses (MOOCs) have increased dramatically recently. As it is stated in [6], researchers are allowed to address problems, which were not accessible few years ago. Recently considering the educational system, Artificial Intelligence (AI) in teaching and learning processes has a surprising evolution. New technologies in education are preferred to help people achieve better educational goals [7]. AI enabled educational tools have gained importance to attract attention also to improve education quality and enhance traditional teaching and learning methods [7]. As stated in different researches, AI technology has developed computer tools for carrying out a number of tasks, simulating the intelligent way of problem solving by humans [8]. The AI technology have also been suggested in the field of special educational needs (SEN) as one of the most valuable applications considered [8]. With the aid

of AI, growth and productivity in many industries such as transportation, communication, commerce, or finance are increased. However, in education field, AI based learning systems need improvements to help the students in classrooms or homes as well as people who required special needs [9, 8]. Considering software developments, instructional software is emerged for learning to individual needs, connects learners together, provides access to digital materials, supports decentralized learning tools and engages students in meaningful ways [10]. Innovations done in technology and multi-disciplinary fields provide new opportunities to change the nature and delivery of teaching without increasing teacher's workloads. It is believed that quality in higher education is increased. One of the fields of AI is artificial neural networks (ANN) are used to understand, differentiate and improve learning strategies of the students by providing detailed educational information to the teachers to improve students' learning strategies [5]. This paper proposed a student performance prediction model that uses machine learning neural network [20]. The proposed work is broadly divided into few blocks such as pre-processing, from the dataset, which help in identifying the information, next is to filter those features from the dataset. Finally, all filtered features train by neural network. The experimental evaluation of the proposed student performance prediction model. All algorithms and utility measures were implemented using the MATLAB tool. The tests were performed on a 2.27 GHz Intel Core i3 machine, equipped with 4 GB of RAM, and running under Windows 7 Professional. The proposed student performance prediction model (SPMM) is compared with the existing methods and found that our model gives better results in all respect.

2. Review of Literature

This section of the research work describes the earlier work done in the field of student performance prediction by the various researchers using different approaches and algorithm for the efficient and accurate analysis of student record.

Utkarsh Verma et al. (2022) uses various machine learning (ML) techniques to predict the student's academic performance using the real data collected (comprising the academic history and personal habits of the students). Furthermore, a comparison of ML techniques on different evaluation metrics has been presented. It will assist the students to keep a track of their academic performance and accordingly, manage their study pattern to help them perform well in future.[19]

J. Dhilipan (2021) proposed a prediction system using their 10th, 12th and previous semester marks. The study is evaluated using Binomial logical regression, Decision tree, and Entropy and KNN classifier. In order to attain their higher score, this framework would assist the student to recognize their final grade and improve their academic conduct.[20].

Juan L. et al. (2020) design effective mechanisms that improve academic results and avoid dropout, among other things. These are benefited by the automation of many processes involved in usual students' activities which handle massive volumes of data collected from software tools for technology-enhanced learning. Thus, analyzing and processing these data carefully can give us useful information about the students' knowledge and the relationship between them and the academic tasks. This information is the source that feeds promising algorithms and methods able to predict students' performance. In this study, almost 70 papers were analyzed to show different modern techniques widely applied for predicting students' performance, together with the objectives they must reach in this field. These techniques and methods, which pertain to the area of Artificial Intelligence, are mainly Machine Learning, Collaborative Filtering, Recommender Systems, and Artificial Neural Networks, among others [11].

Ebenezer et al. (2019) uses student's attendance and various marks scored to evaluate and predict their performance. The model being created will be very useful for all the stakeholders of education system in improving student performance, teaching methodologies and Institution ranking [12].

Diego et al. (2019) proposed a methodology in which the process of data collection and pre-processing is initially carried out, and then in a second stage, the grouping of students with similar patterns of academic performance was carried out. In the next phase, based on the identified patterns, the most appropriate supervised learning algorithm was selected, and then the experimental process was carried out. Finally, the results were presented and analyzed. The results showed the effectiveness of machine learning techniques to predict the performance of students [13]

Hasan et al. (2019) trying to find out student's current status and predict his/her future results. After the outcome, teachers can give him/her proper advice to avoid the poor result and also can groom the student. By finding out the dependencies for final examinations. Which courses he/she should take in the upcoming semester (roles of adviser/teacher). Every year a lot of students lag behind because of lack of proper advice and monitoring. A teacher can't monitor each and every single student at once. If a system can help a Teacher about the students like which student needs which kind of help. Then it will be much helpful for both teachers and student. The aim is helping the student to avoid his/her predicted poor result using Artificial Intelligence. If a student could know what will be his/her result in the future and notify him/her what to do to avoid his/her the bad results by predicting the final examinations mark. This research would be helpful for the

students and teachers with The highest accuracy of 94.88%.[14]

Kumar and Singh (2019) Performance evaluation of students is essential to check the feasibility of improvement. Regular evaluation not only improves the performance of the student but also it helps in understanding where the student is lacking. It takes a lot of manual effort to complete the evaluation process as even one college may contain thousands of students. This paper proposed an automated solution for the performance evaluation of the students using machine learning. A threshold-based segmentation is employed to complete the evaluation procedure over MATLAB simulation tool. The performance of machine learning is evaluated by accuracy and mean square error. The results of our implementation show that the performance of Artificial Neural Network in comparison to Support Vector Machine is better. The Mean Square Error is 5-20% better whereas the Effort Estimation is around 15-27% better. [15]

Rahila Umer et al. (2017) proposed a process mining approach to help in making early predictions to improve students' learning experience in massive open online courses (MOOCs). It investigates the impact of various machine learning techniques in combination with process mining features to measure effectiveness of these techniques. Design/methodology/approach – Student's data (e.g. assessment grades, demographic information) and weekly interaction data based on event logs (e.g. video lecture interaction, solution submission time, time spent weekly) have guided this design. The results show that techniques used in the study are able to make predictions on the performance of students. Overall accuracy (F1-score, area under curve) of machine learning techniques can be improved by integrating process mining features with standard features. Specifically, the use of LR and NB classifiers outperforms other techniques in a statistical significant way. Practical implications – Although MOOCs provide a platform for learning in highly scalable and flexible manner, they are prone to early dropout and low completion rate. This study outlines a data-driven approach to improve students' learning experience and decrease the dropout rate. Social implications – Early predictions based on individual's participation can help educators provide support to students who are struggling in the course. Originality/value – This study outlines the innovative use of process mining techniques in education data mining to help educators gather data-driven insight on student performances in the enrolled courses.[16]

Bydzovska (2016) proposed an approach to predict the students' performance using course characteristics and previous grades. Two different approaches were used. In the first approach, classification and regression were used to predict performance using academic-related data and data about student's social behavior. The findings were significant with the small number of students. In the second approach, collaborative filtering techniques were used to predict the student's performance based on similarity of achievements. Classification algorithms, namely, support vector machine, decision tree, part, IBI, RF, Naive Bayes and rule-based classifier, were used, where support vector machine produced best predictions which were further improved by integrating social behavior data.[17-20]

3. Proposed Methodology

In this section, proposed method is explained step-by-step where different parameters are covered for hiding the sensitive discriminating items while preserving important information which is non sensitive. Whole work is broadly divided into few blocks such as pre-processing, from the dataset, which help in identifying the information, next is to filter those features from the dataset. Finally, all filtered features train by neural network.

3.1 Description

As data, contain different kind of information for the increase of efficiency and understanding of system. So data mining provide approaches for the same. In this work student performance is evaluate its social and academic data. Based on academic patterns features were select which highly affect the student performance by genetic algorithm. Feature set were further process to train the multi-layer neural network that act as flexible framework for predicting the student performance. Results shows that proposed Student Performance Prediction Model (SPPM) has improved comparison parameters from existing student performance prediction models.

Dataset Pre-Processing

This section detailed Student Performance Prediction Model (SPPM). Whole process of developing a prediction framework was divided into three modules first includepre-processing, privacy preserving and pattern generation. Second was feature selection from generated patterns. Third was neural network based learning of selected feature from genetic algorithm. Working of first two modules was shown in fig. 1 and third module was shown in fig.1.

Module 1:

In this module of SPPM preprocessing of raw dataset was performed and this dataset was further processed to improve data privacy from unethical mining. Finally patterns were generated from the dataset based on various feature values of student performance in MTSE dataset.

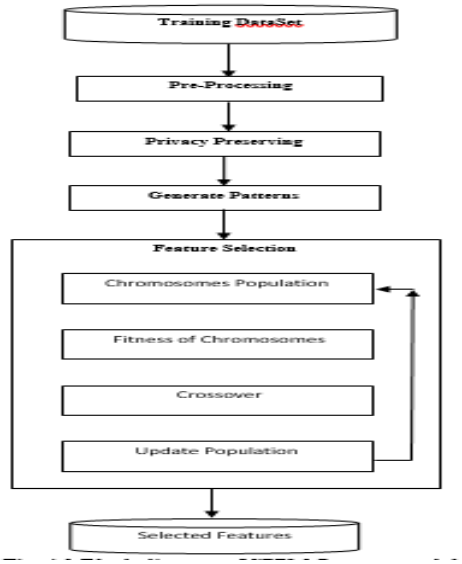

Fig. 1 Block diagram of SPPM first two modules.

As per model requirement input dataset D either training or testing need some cleaning. Some of repetitive information present in MTSE dataset needs to be removed such as school city, school taluka, school district, school address as all these were summarized by pincode. As per pincode locality status were classify into Urban and Rural. As social environmental feature affect learning ability of a student. Transforming the student attempts into subjects class (Maths, Science, General Ability) was also done in this step. As per grade types G each marks were processed to particular grade type.

Module 2:

As student performance were classed into G grades, so this module finds most impacting feature set of students which belongs to a particular G class. Common steps (Generate population, fitness evaluation, and crossover) of genetic algorithm were used to get feasible feature set.

Chromosome Population: Collection of possible feature value set is known as population P_c in the genetic algorithm. Gaussian random number function generates a feature set termed as chromosome C_h, in other words C_h is a subset of F [14]. Population have m number of chromosomes and each chromosome have n number of features for a particular G class.Chromosme were obtained by eq. 1.

$C_h \leftarrow$ Chromosome_generation(F,m,n)----------Eq. 1

Fitness Function: Feasibility of any chromosome was evaluate by fitness function. As per C_h feature subset student matching similar feature were classified into G class [15]. So for student having feature S_f maximum matching chromosome grade F_G was evaluate and if final student F_G is same as training G than fitness value get increment by one.

$$F_G = Max_{Grade(|C_{h,G} \cap S_f|)} \quad \text{Eq. 2}$$

$$F_V = \begin{cases} F_G = S_G \ (F_V increment) \\ F_G \neq S_G (F_V Nochange) \end{cases} \quad \text{Eq. 3}$$

Crossover: Chromosome having high fitness values in the population was consider as the best chromosome in current iteration of genetic algorithm. This chromosome modify other chromosome by replacing few feature values. Ramdom feature were select from best chromosome and it replace other chromosome feature value randomly. This variation gives new chromosome while may increase the prediction accuracy.

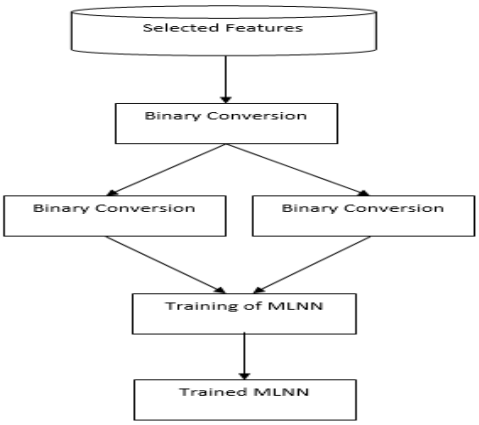

Fig. 2: Selected feature based training of MLNN

Update Population: In this step all new chromosome fitness value were evaluate. Generated fitness value compared with relevant parent chromosome and good fitting chromosome continues in population [16]. After this algorithm moves to fitness function step and continue for t number of iteration.

Selected Feature: Output of this module is a feature set GA_F. After t number of iterations final population fitness value were evaluate and best chromosome act as GA_F.

Module 3

Selected features as per grade were used to train the multilayer neural network as training input vector and desired output vector were need to prepared [17]. Training dataset student features were used to prepare input vector I_v off binary values. Input binary vector have one at matching position of Student feature S_F and genetic algorithm feature GA_F. While non matching features were represent by 0 in the training vector. Desired output of the training vector was D_o. Multilayer neural network model was train by passing I_v, D_o, after sufficient number of epochs neural network get trained. Testing of trained multi layer neural network need features vector of binary format as per student matching feature and give student grade as output.

Proposed SPPM Algorithm

Input: D // D: Raw Dataset

Output: MT_{NN} // Trained Multi Neural Network

Module 1

1. D_P←Pre-Processing(RD) // Processed Dataset
2. D_P←Privacy_Preserving(D_P)
3. P←Single Scan_Counter(D_P)
4. Loop 1:p // n: number of patterns
5. If P>min_supp
6. F[G]←P
7. EndIf
8. EndLoop

Module 2

9. C_h← Chromosome_generation(F,m,n)
10. Loop 1: t //t: number of iteration
11. Loop 1:n
12. $F_{n,G} = Max_{Grade(|C_{h,G} \cap S_f|)}$
13. $F_{n,V} = \begin{cases} F_G = S_G \ (F_V increment) \\ F_G \neq S_G (F_V Nochange) \end{cases}$
14. EndLoop
15. B←Best_Fitness(F_v)
16. Loop 1:n
17. $C_{n,h}$←Crossover(B, $C_{n,h}$)
18. EndLoop
19. Endloop
20. GA_F←Best_Fitness($F_{v,}$,C_h)

Module 3

21. Loop 1:s // s: number of students in training data
22. I_T[s]←Ones(S_F, GA_F)
23. D_o[s]←S_G
24. Endloop
25. MT_{NN}←Mtrain(I_T, D_o)

Above SPPM algorithm gives trained multilayer neural network model, which takes student feature vector and predict grade.

4. Experimental Dataset and Result Analysis

This section presents the experimental evaluation of the proposed student performance prediction model. All algorithms and utility measures were implemented using the MATLAB tool. The tests were performed on a 2.27 GHz Intel Core i3 machine, equipped with 4 GB of RAM, and running under Windows 7 Professional.

Evaluation Parameters

Comparison of academic performance prediction algorithms were done on following parameters.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F\_Measure = \frac{2 * Prcision * Recall}{Precision + Recall}$$

Where TP is true positive, FP is false positive, similarly TN is true negative and FN is false negative. All are counter having 0 initial value. Let us understand for Average Grade. TP increments if a student predicted grade class is average and actual grade class is average. Similarly, TN increments if a student predicted grade class is average and actual grade class is also other than average. In case of FP increments done if a student predicted grade class is other than average and actual grade class is average Similarly FN increments if a student predicted grade class is other then average and actual grade class is also other than average.

Execution Time: Time required generating rules from the input dataset. Total time needed for the finding of patterns, counting of patterns, evaluating rule values.

Dataset

Table 1 Dataset description of MTS exam

Parameters	Values
+Rows	4300
Columns	9
Classes	3
Cities	45
Pincodes	300
School Feature	Outdoor, Presentation, urban/rural, medium, video Lectures, debate, exhibition, tution, homework
Subjects	Science, Social Science, Maths, Mental Ability
Marks Category	Poor, Average, Good

Results Analysis

Table 2 and fig.3 shows precision value of for all grades (Poor, Average, Good). It was obtained that SPPM has increased the precision value by 29.38% as compared to ELA. Use of genetic algorithm in SFLAPF and GIWD has reduced the feature dimension for grade prediction hence prediction get more accurate. It was showed that multi layer neural network has increases the learning accuracy in less number of epochs.

Table 2 Precision based student comparison.

Dataset size	Grade	ELA [18]	SPPM
1000	Poor Average Good	0.1779 0.4181 0.2643	0.4017 0.4861 0.375
1500	Poor Average Good	0.1787 0.4368 0.2702	0.3834 0.4981 0.3573
2000	Poor Average Good	0.1826 0.4302 0.2721	0.3952 0.5000 0.3625
2500	Poor Average Good	0.2444 0.4289 0.2645	0.4513 0.4852 0.3577

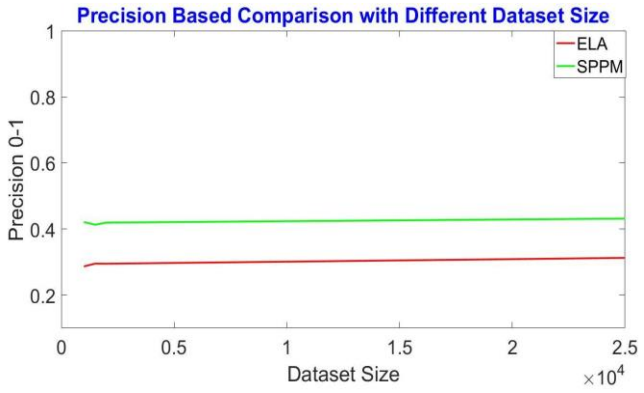

Fig. 3 precision value based comparison of student performance.

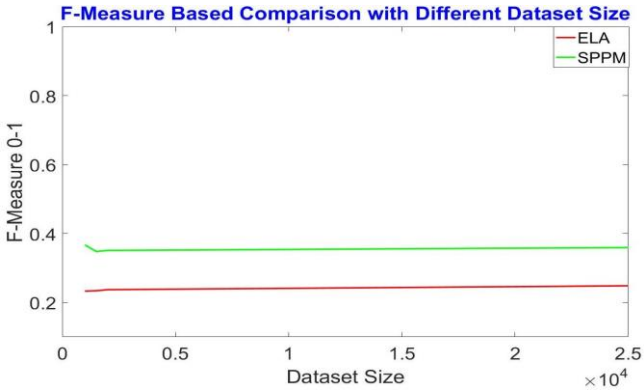

Fig. 4 F-measure value based comparison of student performance

Table 4 F-Measure based student comparison

Dataset size	Grade	ELA [18]	SPPM
1000	Poor	0.1495	0.3968
	Average	0.3405	0.4263
	Good	0.209	0.2780
1500	Poor	0.1459	0.3321
	Average	0.3436	0.4312
	Good	0.2127	0.2806
2000	Poor	0.1515	0.4062
	Average	0.3462	0.3930
	Good	0.2139	0.2537
2500	Poor	0.1981	0.3747
	Average	0.3380	0.4246
	Good	0.2092	0.2788

Table 4 and fig. 4, shows that F-measure, accuracy parameter of student performance prediction was high at different dataset size. Results shows that SPPM has increases the grade class prediction value in all set of poor, average, good segment.

5. Conclusion

Data mining is one of the important fields in which researchers are working today. College and university students are any country's untapped resource. To raise or lower a group of students' average quality, education depends on a variety of elements. In order to extract various feature patterns from the data, this article utilised the single scan counter data analytics model to the data. For the purpose of gathering and archiving pertinent student feature data according to grade levels, Paper has created an extensible framework. Crossover operation was employed for this genetic strategy. In a subsequent publication, a multilayer neural network (MLNN) that was used to create a learning model was trained. During testing, the MLNN identified pertinent parameters based on chosen student attributes.

On a genuine dataset, experimental work was done, and the outcomes were compared on several parameters. It was discovered that SPPM had a prediction precision value that was 29.38% higher than the ELA model. Additionally, it was discovered that the prediction performance of the model with neural network SPPM was superior to the model whose F-measure value had grown by 33.16% compared to ELA. Future perturbation methods could be improved and made new.

- This work have not used a prior rule for grouping which give better results.
- Scholar can go with utilization of the segmenting scheme may give fruitful results.

As individual privacy is, measure concerns so proper steps required for maintaining their privacy without high loss of data.

References

[1] J. A. Sanguesa, V. Torres-Sanz, P. Garrido, F. J. Martinez, and J. M. Marquez-Barja, “A review on electric vehicles: Technologies and challenges,” Smart Cities, vol. 4, no. 1, pp. 372–404, 2021, doi: 10.3390/smartcities4010022.

[2] G. N. Kumar, T. A. Kumar, and G. V. Madhav, “Bridgeless Cuk Converter with Fuzzy Controller for Power Quality Enhancement of PV Integrated EV Charger Turkish Journal of Computer and Mathematics Education Research Article,” Comput. Math. Educ. Turkish, vol. 12, no. 12, pp. 2046–2063, 2021.

[3] M. P. Jati, “A Power factor corrector using interleaved boost fuzzy-logic converter: design, analysis, and implementation,” J. Eng. Appl. Technol., vol. 2, no. 1, pp. 23–35, 2021, doi: 10.21831/jeatech.v2i1.39473.

[4] A. K. Singh, A. K. Mishra, K. K. Gupta, and T. Kim, “Comprehensive review of nonisolated bridgeless power factor converter topologies,” IET Circuits, Devices Syst., vol. 15, no. 3, pp. 197–208, 2021, doi: 10.1049/cds2.12046.

[5] L. Corradini, G. Spiazzi, and Z. Zhang, “Special Issue on Topologies, Modeling Methodologies, and Control Techniques for High-Frequency Power Conversion,” IEEE J. Emerg. Sel. Top. Power Electron., vol. 8, no. 3, pp. 1967–1970, 2020, doi: 10.1109/JESTPE.2020.3000529.

[6] V. Kalyanasundaram, S. George Fernandez, K. Vijayakumar, and S. Vidyasagar, "A two stage battery charger for EV charging applications," Indones. J. Electr. Eng. Comput. Sci., vol. 19, no. 2, pp. 593–599, 2020, doi: 10.11591/ijeecs.v19.i2.pp593-599.

[7] B. Dimitrov, K. Hayatleh, S. Barker, G. Collier, S. Sharkh, and A. Cruden, "A buck-boost transformerless DC–DC converter based on IGBT modules for fast charge of electric vehicles," Electron., vol. 9, no. 3, 2020, doi: 10.3390/electronics9030397.

[8] M. Brenna, F. Foiadelli, C. Leone, and M. Longo, "Electric Vehicles Charging Technology Review and Optimal Size Estimation," J. Electr. Eng. Technol., vol. 15, no. 6, pp. 2539–2552, 2020, doi: 10.1007/s42835-020-00547-x.

[9] M. Bharathidasan and V. Indragandhi, "Review of Power Factor Correction (PFC) AC/DC-DC Power Electronic Converters for Electric Vehicle Applications," IOP Conf. Ser. Mater. Sci. Eng., vol. 906, no. 1, 2020, doi: 10.1088/1757-899X/906/1/012006.

[10] S. Vijai, M. Amutha, M. I. A, G. Priya, and S. Vinoth, "Efficient PFC with Switched Inductor DC-DC Converter for Battery Charging Application," vol. 7, no. 02, pp. 1–6, 2019.

[11] R. Mahalakshmi and R. Seyezhai, "Review of Integrated Power Factor Correction (PFC) Boost converter topologies for Telecommunication systems," Int. J. Adv. Res. Basic Eng. Sci. Technol., vol. 2, no. 11, pp. 13–23, 2016.

[12] S. A. Alim and S. Islam, "Open Access A Proficient AC / DC Converter with Power Factor Correction American Journal of Engineering Research (AJER)," no. 8, pp. 233–238, 2016.

[13] A. Sharma, "Critical review on electric and plug in hybrid electric vehicles," no. December 2015, [Online]. Available: https://www.researchgate.net/publication/326059591

[14] D. De Castro Pereira, M. R. Da Silva, E. M. Silva, and F. L. Tofoli, "Comprehensive review of high power factor AC-DC boost converters for PFC applications," Int. J. Electron., vol. 102, no. 8, pp. 1361–1381, 2015, doi: 10.1080/00207217.2014.981871.

[15] Vaseem Naiyer, Jitendra Sheetlani, Harsh Pratap Singh, "Software Quality Prediction Using Machine Learning Application", Smart Intelligent Computing and Applications, 2020, Springer.

[16] Abdul Rasool Md, et al. "Data Mining Approaches to Identify Spontaneous Homeopathic Syndrome Treatment", Annals of the Romanian Society for Cell Biology,2021. Pp 3275-3286.

[17] K Nagi Reddy et al. "Nb-Uvb Phototherapy treatments for Short-Term Adverse Events Using Dataminig Technique", Annals of the Romanian Society for Cell Biology, 2021, pp 3475-3487.

[18] Pasha and Singh, "A Novel Model Proposal Using Association Rule Based Data Mining Techniques for Indian Stock Market Analysis", Annals of the Romanian Society for Cell Biology, 2021, pp 9394-9399.

[19] Utkarsh Verma; Chetna Garg; Megha Bhushan; Piyush Samant; Ashok Kumar; Arun Negi, "Prediction of students' academic performance using Machine Learning Techniques", International Mobile and Embedded Technology Conference (MECON), 2022, In proceeding of IEEExplore.

[20] J. Dhilipan, N.Vijayalakshmi, S.Suriya, Arockiya Christopher, "Prediction of Students Performance using Machine learning", IOP Conf. Series: Materials Science and Engineering 1055 (2021) 012122.