

Filter-Wrapper Feature Selection with Improved Deep Learning for Education Data Mining

Dr. M. Rajeshwari,

Assistant Professor, Department of Computer Applications, Sri Ramakrishna College of Arts & Science (Autonomous),
Coimbatore – 641 006

Dr.R.Anushiya,

Assistant Professor, Department of Computer Science with Data Analytics, Sri Ramakrishna College of Arts and Science,
Coimbatore - 641 006

Abstract - The quality of learning models has a huge impact in deriving useful information from data. For this scenario, preprocessing handles inconsistencies in data to improve the model's accuracy. The foremost major steps in preprocessing are data Quality Assessment, Cleaning, Transformation, and Feature Selection prior to Machine Learning Algorithms. The first phase of this research work comprise of Feature Selection to have relevant features that gives the information about significant features also gains the accuracy and Supervised Classification in order to categorize the learners based on their performance. This work proposes Filter-Wrapper Feature Selection method. The method starts with assigning weights using Filter method to each Feature based on Symmetrical Uncertainty and creates a subset of Features with weights. Instead of applying all attributes to the Wrapper feature subset method, the attributes with higher weights which are obtain from Filter method are only applied in order to minimize the computation complexity and to improve accuracy. Generally, the wrapper Feature method includes a learning algorithm to have better feature as well as classifies the dataset. For the learning algorithm, this paper applies Deep Multi FeedForward Neural Network with AdaDelta Optimizer, neural network parameters. The performance is assessed in Rapidminer tool using the Classification evaluation measures namely Accuracy, Weighted Mean Recall (WMR), and Weighted Mean Precision(WMP).

Keywords – *Deep Multi FeedForward Neural Network, Education Data Mining, Symmetrical Uncertainty.*

1. INTRODUCTION

Data mining (DM) in other terms Knowledge Discovery in Databases (KDD) as it derives knowledge from data encapsulates Machine Learning (ML) and statistical analysis to make patterns and precious information. [1] This is rapidly accelerated over the last decades and can be automated with ML and Artificial Intelligence (AI) that aids in speed up the processes.

Experimenting the Data in untraditional ways generate outcomes which was quite remarkable and advantageous. There begins the management of Data Mining with the progression of warehouse and database. This advancement allows all disciplines to store enormous data that makes analyzing in efficient and quick manner. Data mining gently comprise of five vital steps as follows: Setting the Objectives, Selection of Data from the appropriate Source Domain, Preparing the Data as per the requirements, Making the Model with Data Mining using the methods Regression, Clustering, Association Mining, Sequential Analysis, Classification etc., Revealing Pattern, and Evaluating

the outcomes and in final presenting the knowledge from the derived patterns.

(a) Set the Objectives - The problem should be defined extensively in order to set the goal that helps to derive the data questions and parameters of the project. The objectives should reveal the motivation clearly. The primary objective is to extract applicable perception from basic data, enabling the organizations to make knowledgeable choices and predictions.

(b) Data Selection - The data should be opt for answering the pertinent questions. It might be in the form of numeric, categorical, ordinal, nominal based on the domain. The two primary types are quantitative and qualitative. The optimum data should always guarantee validity and reliability. It encompass subtasks such as choosing the opt source, defining the variables of the data, constructing Meta data and selecting specific samples. Now-a-days, data selection tool such as HPXML is available to select and to get the the requirement of it.

(c) Data Preparation - The relevant data should be free from missing values, duplicates, noise, and outliers. Based on the type of analysis, the dimension of the data can also be reduced that can aid in speeding the process. The final accurateness highly depends on the data being used and so it should be transformed accordingly because data can arrived in any sources, forms that might be structured or unstructured.

(d) Model building with Mining, Pattern Extraction - The raw data with labeled or unlabeled may contain any kind of relationships, such as correlations, groups, categories, sequential patterns, or association and in some cases the deviations in the data are more interesting which is referred as outliers that highlights certain data. The predictive models derive future trends with respective to present one. Depending on the

analysis, the data mining methods are building to extract the interesting pattern.

2. EDUCATION DATA MINING (EDM)

The information generated from educational settings intended to design and enrich methods for examining the data with multiple meaningful educational hierarchy. [2] It discovers new insights to get the knowledge regarding how the people learn in such settings. Thus, it is contributed to learning sciences and educational psychology. It refers to tools, techniques, and research that designed to automate the extraction of meaningful pattern from large or small repositories of education data which are related to learning activities. Most commonly the data are fine grained, extensive and precise. Figure 1, shows the EDM with its types, tasks, and methods.

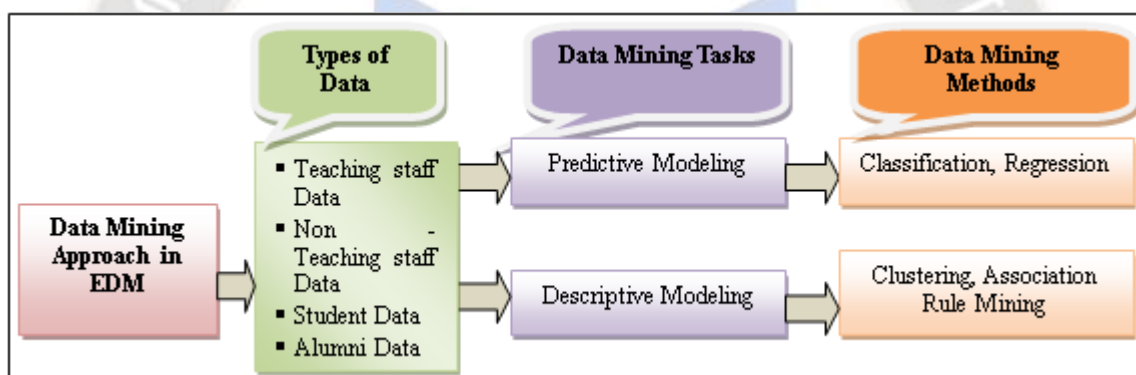


Figure 1. Education Data Mining

2.1 Different Types of Learning Environments

The learning environment varies as per the learner requirements; it includes the intellectual, social, and emotional feeling all of which will affect learning. In addition, the demographics, accessibility, and the way peers interact will also play a key role. Below are the important types of learning environments namely Traditional, Online, and Education Software classes. [3] The data in this environment are applied with DM methods and the final pattern presents the knowledge to the educators regarding the learners and this in turn useful for the learners to have the recommended service on the studies.

Classroom Learning – It is a traditional mode from kinder garden, schools-senior schools, college/universities and most of it are depends on traditional learning space environment with tutor and learner need to be in presence in the room. It is

advantageous in the way it builds a relationship between the facilitator and the learner hence it makes mentor-mentee relationship and promotes networking. However, in case of pandemic situation, this might consider as a potential zone of risk. Now a days, the distance, duration of travel time to the institution in the city environment becoming very earlier even for kids leads to complication mentally, physically. In contrary, a digital classroom have physical classroom with technologies for education such as laptops, smart whiteboards or Virtual Reality headsets. This type of environment records student information such as ID, course selection, attendance, assignment marks, internal marks, practical, model exam, total marks, grade.

Online Learning– Most commonly, the teachers and students can present course materials, interact, and engage them and work in groups together. It enables distance education for learners. It offers customized

learning experience. In Asynchronous online course, [4] the learners are provided with materials assignments, a period for the course. In this scenario, the interaction will be via blogs, wikis and discussion boards. It is suitable and effective for the learners who have time constraints or busy schedules. In contrast, synchronous course necessitates instructor and the learners to work in online and so regular attendance is recorded, thus it enables the students to be a part of a curriculum from distance. The type of attributes to describe each student include family size, location, device type, internet stability, load shedding, CGPA, class duration, status level economically, sports interest, study, social media & sleep spending time, clearing doubts, interaction level.

Education software –Implies the practice of using computer applications for enhancing the learning process to serve the need of students, teachers, and administrators. The programs provide interactive and personalized learning experiences as it works on various platforms, devices, makes the learning process more accessible, effective and flexible. The features include enhanced engagement, personalized learning, media integration, advanced interaction. The categories include course material, classroom needs, assessments, reference, platforms, training & tertiary education. The attributes in this type of data are: the organization providing the course, title of the course, skill covered in the course, rating, difficulty level, type of certificate, duration, and credit eligibility.

2.2 Applications of EDM

Prediction of Students' Learning Assessment – [5] Incorporating learner's characteristics that include student academic assessments quizzes, and interactions to create student models which is then used to predict the learning behaviour of each student effectively. The student experience and their satisfaction regarding learning objects are measured.

Improving the Education Domain – Through the existing methods with the improvements the domain is enriched. The current educational content engages the learners and determine the instructions to support the curriculum style. The domain is improved by the analysis of feedback data from various stakeholders that includes students, parents, and faculty members.

Assessing the Effects of Educational Support – The educational support system includes the learning objects used for education system. This includes the learning objects assessment for improving the system of study.

The institutions can optimize resource allocation by analyzing the data such as classroom utilization, workload, and number of student enrollment.

Optimizing Course Offerings – The universities can optimize their course by identifying courses in high demand, which implies most liked by learners and in contrary, the courses that are not selected which implies that should be improved.

Improving Teaching Effectiveness, Institutional Performance - Provides insights into instructional practices and student learning outcomes. By analyzing on student engagement data, the institutions can identify areas where the teaching can be improved and make development courses for faculty.

Future Perspectives of Education Data Analysis - With the advancements in technology and an increasing recognition in EDM, the future of data analysis in education looks promising. It will continue to outline the way of educators teaching, students learning, and formulating educational policies and so provide deeper insights into learning. The instructors tailor their teaching methods more precisely that leads in improved learning outcomes.

The rest of the section is organized as follows, section 1 has the introduction part, section 2 discuss about Education Data Mining, section three lists the existing literatures, section 4 explains the proposed research methodology, section 5 has the results part with performance evaluation and section 6 concludes the research work.

3. LITERATURE REVIEW

Raju *et al* [6] did a report based on the survey on Education Data to draw vital information and to reveal the correlation among the variables. This survey discussed the various approaches that are grouped under two categories namely, Super-Unsupervised indulged to analyze student behaviour in exams. Seven types of Classification and Clustering oriented methods are discussed. From the overall survey, it can be understood the performance and the information about the Learning methods discussed for EDM.

Syed Amin Ullah *et al* [8] attempted a comparison to predict a student's academic with various classification techniques. For the experiments, a five years MBBS dataset obtained from the institutions. After preprocessing, 10 fold cross validation with the classifiers was implemented in WEKA tool. Among the

11 classifiers, Hoeffding Tree and Naive Bayes gave highest accuracy with 59.45%. Finally, from this comparative analysis, it is noted that accuracy level was very less even though the processing time is acceptable.

Thaer Thaher *et al* [7] introduced an efficient student's performance model using MLP approach. Three public educational datasets were collected from the repositories. For optimization technique the approaches such as stochastic gradient, Adam, Adaptive gradient Quasi Newton were assessed with the meta heuristic method namely PSO, Harris Hawk and Grey Wolf. In Neural, tanH activation and MSE loss function was utilized. From the overall comparison the method MLP_Adam with random Oversampling gave high AUC value with 0.91.

Dinesh Kumar *et al* [9] examined the academic information in educational data. The work presented a hybrid approach after feature subset Pearson correlation with 6 features. Initially, the classifiers were executed separately and then made hybrid to assess the accuracy. For classification, RB Network, MLP Network, DT, and RF with 10 fold cross validation were utilized. The UCI repository student data was experimented in WEKA. The hybrid Network achieved 75.6% and hybrid Trees obtained 76.4% which are higher than the individual methods.

Jaikumar M. Patil *et al* [10] designed a work in order to inspire educational institutions in different aspects of their assessment. Five stages that includes problem statement, data collection, preprocessing, classification and finally testing and prediction analysis were made. The experiments was made in WEKA tool in synthetic datasets. Data sampling, balancing, normalization was done in preprocessing. Decision tree, Kernel, Networks, Probabilistic based classifiers were taken for the implementation and found SVM with kernel has 93% accuracy.

Mustafa Yagci *et al* [11] presented a method for the early assessment of high-risk failure of students. This work proposed a model to predict the final-exam grades in prior with the previous marks. The dataset was taken from University in Turkey of 1854 students with mid-term exam grades in the Department. Seven procedures with 10-fold validation were evaluated. Only three variables namely grades, faculty and department information was considered for the implementation. The experiment was performed in Orange. From the results, it was noted the Random Forest and Neural-Network

achieved 76.6% accuracy. However, the accuracy was less that can be improved.

Kamil Yurtkan *et al* [12] proposed an approach to predict student using Feed-Forward. For Deep Network, two hidden layers were used. After the variance based feature selection, the classifier with 10-fold validation was implemented in MATLAB. The results showed that the accuracy was high with 93.1% for the whole dataset rather than the feature subset data that has 91.4%. Hence it is concluded, the proposed Deep Feed Forward Neural Network should be adjusted repeatedly and prompt feature selection method should be used to retain the accuracy of the learning process.

Noora Shawareb *et al* [13] focused on the data in LMS specifically Moodle, that has 20 attributes. The work comprise of four stages namely finding the descriptive attributes, fixing the target attribute, applying machine learning and assessing the performance. The collected data underwent pre-processing and then different machine learning techniques were applied. The experiments were carried out in MATLAB 2018b. With the random shuffling, the data was parted as 80:20. The outcome showed that the MLPNNs with 10 neurons in the hidden layer achieved accuracy of 93%.

Zara Ersozlu *et al* [14] investigated machine-learning approaches for educational data. This paper did a systematic review by examining 77 publications. In the result section, the common keywords and publication trends, application domain such as predicting academics, dropouts, learning process, teaching quality, algorithms used for EDM in the last decade were explained. The analysis showed that 88% of the publications used supervised approaches and the semi-supervised were less in common.

Chitra Jalota *et al* [15] analyzed various techniques that are used for predicting academic. The work discussed the classification methods under several categories namely statistical based, Correlation, Regression, Bayesian, Distance, Trees, Network and Rule based. For the experimentation, dataset gathered from LMS was applied on Weka. The features were categorized as demographical, educational and psychological with 163 instances. The target has three grades and five algorithms were applied. The accuracy of the classifiers showed MLP achieved high values with 76.07% and 124 correctly classified instances, sixteen attributes followed by SVM.

4. RESEARCH METHODOLOGY

Filter-Wrapper Feature selection using Improved Deep Learning with FeedForward Neural Network (FW-IDL)

This research work aims to build Feature Selection method with Machine learning algorithm. In order to handle the challenges such as computational complexity, high processing time in Wrapper method, Filter method is used in prior to assign weight to the Features. Then the higher weight attribute is applied for wrapper method which includes a learning algorithm in default, here the proposed method improved Deep Learning method (IDL) with FeedForward Neural Network that leads to a single flow of work which produces the output with the relevant feature set and classified results.

Filter Method using Symmetrical Uncertainty (SU)

It is used to have the non-linear relationship among variables which is crucial in Machine Learning and Artificial Intelligence. This method formulated using Shannon Entropy and Information Gain. For the set of two variables (A, B) { Where, $A=\{a_1, a_2, \dots, a_n\}$ and $B=\{b_1, b_2, \dots, b_n\}$ } the SU is calculated by,

$$SU(A, B) = 2 * \left[\frac{IG(A, B)}{E(A) + E(B)} \right] \quad (4.1)$$

Where,

$$IG(A, B) = E(A) - E(A|B) \quad (4.2)$$

$$E(A) = - \sum_i P(a_i) \log_2(P(a_i)) \quad (4.3)$$

$$E(B) = - \sum_j P(b_j) \log_2(P(b_j)) \quad (4.4)$$

$$E(A|B)$$

$$= - \sum_j P(b_j) \sum_i P(a_i | b_j) \log_2(P(a_i | b_j)) \quad (4.5)$$

The value $P(a_i)$ and $P(b_j)$ is defined as the probability of the occurrences of the i^{th} value in the variable a and j^{th} value value in the variable b respectively. $E(A|B)$ is the conditional entropy.

Let there are features $V_1, V_2, \dots, V_n$, with the class T , the features are assigned weights using SU as in equation (4.1) and the weights are arranged in descending order. Here, instead of user assigned threshold value (weight), the higher weight features are made as initial candidate set (subset) for the Wrapper method. Hence, this method alternates the usual procedure of taking the full feature set thus minimize the computation complexity of the existing Wrapper method.

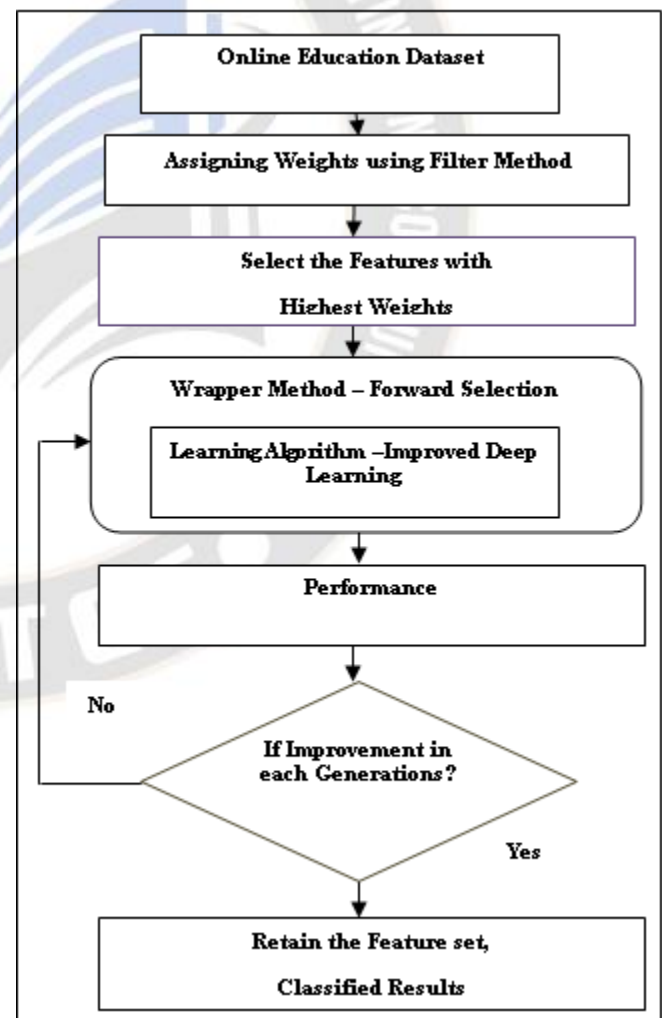


Figure 2. Proposed Flow of Work FW-IDL

Figure 2, shows the flow of the proposed work FW-IDL. It starts with filter feature selection, based on the results higher weighted attributes are selected and then the wrapper method is applied to have most relevant attributes and classification results also obtained using the same as the learning algorithm is combined.

Wrapper Forward Feature Selection, Improved Deep Learning Classification

Wrapper Forward Selection - The learning algorithm creates a subset of features starts from one to all or all to one, compares the results of every execution analysis and then confirms the subset, this leads multiple execution, longer processing time, but accuracy will be high. The higher accuracy feature subset is set as the model.

Deep Learning - A Hierarchy of layers is adopted to transform the input to represent the output. The extra layers used to analyze complicated features of the input. It comprise of neurons, weights, bias, synapses, and functions. It is Supervised that can classify or predict data. The proposed network is FeedForward Neural Network.

Deep FeedForward Neural Network – The proposed work has 3 hidden layers, 25 neurons each. The input, output neurons are based on the input feature subset and the label/class of the dataset. Figure 3 shows the Deep FeedForward Neural Network with three hidden layers.

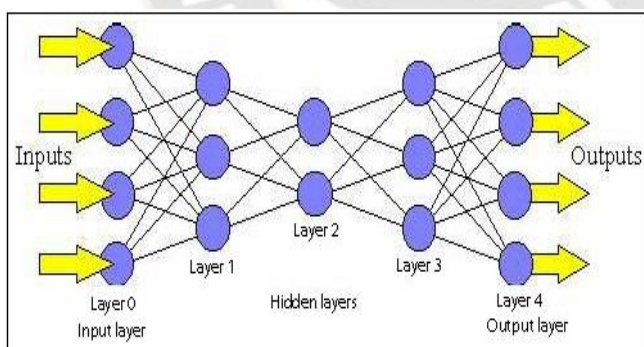


Figure 3. Deep FeedForward Neural Network

The bias initialized with 0.0, Epoch 100, and the weight with $(0 - \sigma)$ using 'He Normal' method with no. of inputs I by,

$$= \text{Sqrt}\left(\frac{2}{\text{Inputs}}\right) \quad \sigma$$

– (4.6)

The proposed Activation function is Exponential Linear Unit (ELU) which base form is Rectifier Linear Unit (ReLU). Hence, it is also referred as Exponential Rectifier Unit. The alpha is added with base ReLU that makes the function more smoothness when inputs are non-positive.

$$E(I) = \begin{cases} I & \text{if } I > 0 \\ \alpha(e^I - 1) & \text{if } I \leq 0 \end{cases}$$

– (4.7)

Multiclass Cross Entropy loss function falls between the range $(0 - \sigma)$ is used as the function to estimate the error with the output O_i by using,

$$CL = - \sum_{i=1}^n O_i \cdot \log \hat{O}_i$$

– (4.8)

Stochastic Gradient Descent is used to update the model, starts from the aim of reducing the loss of the function, the parameter weight is updated for each training instances one by one, handles large dataset efficiently. The process starts with shuffling the training set, iterate over the each sample for computing the gradient of the loss function scaled by learning rate, momentum to reduce the error. The convergence criteria for this iterations is to get the minimum difference in the loss functions between this gradient iterations.

$$G := G - \eta \nabla Q_i(G)$$

– (4.9)

AdaDelta Optimizer overcomes two issues such as initialization of starting learning rate and decaying learning rate with two state variables namely t_p that have the leaky mean value of the 2nd moment of the gradient, and u_p . The stochastic can be enhanced using AdaDelta with d_p equals $\nabla Q_i(G)$, the gradient at the iteration period t is calculated as,

$$t_p = \rho t_{p-1} + (1 - \rho) d_p^2$$

– (4.10)

$$= u_{p-1} - d'_p - (4.11)$$

Where,

$$= \frac{\sqrt{\Delta u_{p-1} + \epsilon}}{\sqrt{t_p + \epsilon}} \cdot d_p - (4.12)$$

$$= \rho \Delta u_{p-1} + (1 - \rho) d_p^2 - (4.13)$$

Multinomial Distribution Function comprise of k possible results $(L_1, L_2, L_3, \dots, L_k)$ with associated probabilities $(p_1, p_2, p_3, \dots, p_k)$ in where $\sum p_i = 1$, as one outcome must certainly occur. For repeated trials of n , may t_i indicate the no. of times for the outcome L_i happens, such that $0 \leq t_i \leq n$ and $\sum t_i = n$. It is written as,

$$P(L_1 = l_1, L_2 = l_2, \dots, L_k = l_k) = \frac{n!}{l_1! l_2! \dots l_k!} \cdot p_1^{l_1} p_2^{l_2} \dots p_k^{l_k} \quad (4.14)$$

Algorithm for FW-IDL

Input	Education Sector Dataset
Output	Feature Subset, Classified Instances.
Step 1:	Preprocessing – Filtering the Missing Values
Step 2:	Load the Dataset
Step 4:	Assign Weights to each m Attributes using Symmetrical Uncertainty Sort the Dataset based on the Weights in Descending order.
Step 5:	Apply Weight Guided Wrapper Feature set using Deep FeedForward Neural Network using, Deep FeedForward Neural Network with appropriate Network Parameters, Functions Evaluate the model performance of each K , Store the K performance, Average the performance of K values
Step 7:	Stop the Process.

u_p

d'_p

Δu_p

Advantages of FW-IDL

- ⇒ The proposed method designs an improved Feature Selection with Classification in a single window that comprise of 'Filter-Wrapper Feature Selection with Classifier'.
- ⇒ Deep FeedForward Neural Network is analyzed for variant number of Hidden Layers, Neurons and Dropouts and provides a clarification to the uncertainty of network architecture.
- ⇒ The proposed method increases the accuracy and minimizes the computation complexity.

5. RESULTS AND DISCUSSION

This section consists of Dataset description, Proposed Methodology, Results and Performance evaluation using the suitable multiclass classification measures.

5.1 Dataset Description

The dataset named Online Education Dataset is taken from Kaggle repository (www.kaggle.com/datasets/sujaradha/online-education-system-review). This data collection is about to analyze the impact of online education. It comprises of 23 attributes and 1033 instances. Table 1, shows the list of attributes with description and data types.

Table I. Online Education Dataset

No	Attributes	Attribute Description	Data Type
1	Gender	Male, Female	Nominal
2	Home Location	Rural, Urban	Nominal
3	Level of Education	Post Graduate, School, Under Graduate	Nominal
4	Age	Years	Numeric
5	Number of Subjects	1-20	Numeric
6	Device type used to attend class	Desktop, Laptop, Mobile	Nominal
7	Economic Status	Middle class, Poor, Rich	Nominal
8	Family size	Numbers	Numeric
9	Internet facility	Range from Very	Nominal

	in your locality	bad to Very Good	
10	Are you involved in any sports	Yes/No	Nominal
11	Do elderly people monitor you	Yes/No	Nominal
12	Study Time	Hours	Numeric
13	Sleep Time	Hours	Numeric
14	Time spent on social media	Hours	Numeric
15	Interested in Gaming	Yes/No	Nominal
16	Have separate room for studying	Yes/No	Nominal
17	Engaged in group studies	Yes/No	Nominal
18	Average marks scored before in traditional classroom	Range	Numeric
19	Your interaction in online mode	Range from Very bad to Very Good	Nominal
20	Clearing doubts with faculties in online mode	Range from Very bad to Very Good	Nominal
21	Interested in	Practical, Theory, Both	Nominal
22	Performance in online	Range from Very bad to Very Good	Nominal
23	Level of Satisfaction-Target	Average, Good, Bad	Nominal

5.2 Implementation Tool – Rapidminer

A Data Science platform written in Java programming created to analyze the organization's data, in 2006, it was Rapid-I then rebranded as RapidMiner on 2013 that make use of client/server model. This provides Data Loading, Transformation, Preprocessing, Mining & and Visualization. It has GUI provided to design and execute workflows. The workflow is

modeled in such a way that the output of each operator can be formed as the input of the next one.

5.3 Experimental Performance Analysis

The OED dataset has the performance of the student in online and offline classes separately. The actual dataset has 3 labels namely Average, Good, and Bad that is based on only the offline performance. The proposed method finds overall performance of the student that includes offline performance and online performance is calculated and based on that a label is created. The students were labeled with 5 categories based on their performance namely 1- Excellent, 2-Good, 3-Average, 4-Below Average, 5-Least that is shown in Figure 4.

id	Label	Interested in?	Average ma...	Gender
id_1012	5	3	4	Male
id_1013	3	2	4	Male
id_1014	3	2	4	Male
id_1015	3	1	4	Male
id_1016	5	1	4	Male
id_1017	4	2	4	Male
id_1018	3	2	4	Female
id_1019	5	2	4	Female
id_1020	4	2	6	Male
id_1021	4	2	6	Female
id_1022	4	2	6	Female
id_1023	4	1	6	Male
id_1024	5	3	6	Male
id_1025	4	2	6	Male

Figure 4. Runtime Screen of Proposed Label with 5 categories

Table 2, shows the accuracy level for the proposed method with variant number of hidden layers that constitute the Deep Learning, number of neurons. From the Table 2, it is observed that for layer 3, neurons 25, and the proposed method FW-IDL gives high accuracy as 95.8%

Table II. Accuracy for the Proposed Method FW-IDL for variant

Hidden Layers, Neurons		
Layers	Neurons	Accuracy in %
2	25	88.7
	50	83.4
	75	84.4
3	25	95.8
	50	90.2
	75	86.8
5	25	83.4
	50	84.1
	75	78.4
10	25	82.1
	50	79.2
	75	72.9
12	25	74.5
	50	70.2
	75	70.7

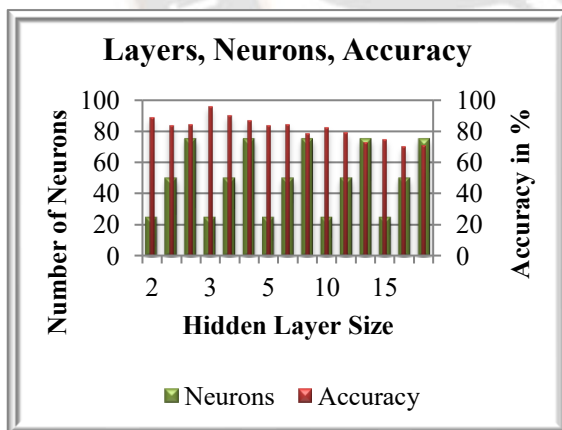


Figure 5. Accuracy for the Proposed Method FW-IDL for variant

Hidden Layers, Neurons

In Table 3, the performance of the proposed method FW-IDL is assessed with existing benchmark methods using the metrics namely Accuracy, WMR(Weighted Mean Precision), WMP(Weighted Mean Precision).

Table III. Performance Analysis of FW-IDL

Methods	(Values in Percentage)		
	Accuracy	WMR	WMP
MLP_ADAM	85.7	78.1	74.9
SVM	91.1	87.2	83.5
FW-IDL	95.8	91.9	89.5

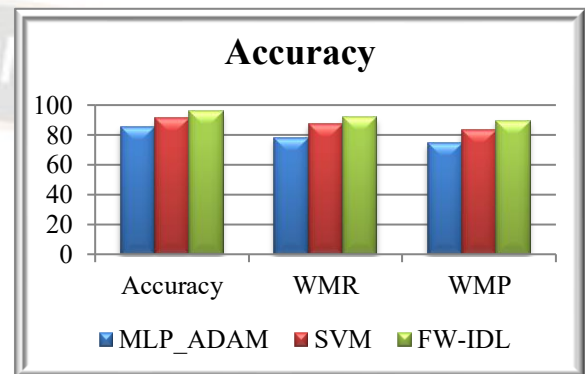


Figure 6. Performance Analysis of FW-IDL

Figure 6, shows the performance of the proposed methods with two recent benchmark data mining methods and it is confirmed that proposed method achieves 10.1% high accuracy than MLP_ADAM and 3.7% higher accuracy than SVM. The WMR, and WMP values also high for the proposed method.

6. CONCLUSION

This research work develops an Educational Data Mining using Improved Deep Learning with Filter-Wrapper Feature selection method FW-IDL. This method minimizes the computation complexity of the Wrapper method as well as gives high accuracy. The dataset OED is taken for analysis. This research analyzes OED dataset in the following aspect: the cumulative performance of the student includes online, offline is find out with the performance categories 1-Excellent, 2-Good, 3-Average, 4-Below Average, and 5-Least. Deep FeedForward Neural Network is applied as a learning algorithm. The proposed method gives 95.8% Accuracy for OED dataset.

In future, the work may be extending for the different analysis in education dataset in the way it may include sport quota analysis. In addition variant Deep Learning method can be utilized.

REFERENCES

- [1] Arati Shah, "Handbook on Data Mining", Redshine Publication, ISBN-10:9394727655, 2021.
- [2] Abdul Aleem, Manoj Madhava Gore, "Educational Data Mining Methods: A Survey", IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT), IEEE Publication, 2020, pp.182-188
- [3] Agung Triayudi, Rima Tamara Aldisa, S. Sumiati, "New Framework of Educational Data Mining to Predict Student Learning Performance", volume: 15, number 1, Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), 2024, pp. 115-132.
- [4] Balqis Al Breiki, Nazar Zaki, Elfadil A. Mohamed, "Using Educational Data Mining Techniques to Predict Student Performance", International Conference on Electrical and Computing Technologies and Applications (ICECTA), IEEE Publication, 2019.
- [5] Chaka Chaka, "Educational data mining, student academic performance prediction, prediction methods, algorithms and tools: an overview of reviews", Journal of E-Learning and Knowledge Society, Vol. 18, No. 2, 2022, pp. 58-69
- [6] Raj Mani Shukla, Shamik Sengupta, Scalable and Robust Outlier Detector using Hierarchical Clustering and Long Short-Term Memory (LSTM) Neural Network for the Internet of Things, Internet of Things, 2020.
- [7] Syed Amin Ullah, Mohib Ullah, Rafiullah Khan, Kamran Ullah, Yasir Ahmed, Atta Ur Rehman, "Educational Data Mining: A Comparative Study to Predict Student Academic Performance Using Deep Neural Network and Other Machine Learning Techniques", 1st International Conference on Computing Technologies, Tools and Applications, ICTAPP23, Institute of Computer Sciences and Information Technology (ICS/IT) 2023.
- [8] Thaer Thaher, Rashid Jayousi, "Prediction of Student's Academic Performance using Feedforward Neural Network Augmented with Stochastic Trainers", 14th International Conference on Application of Information and Communication Technologies (AICT), IEEE Publication, 2020.
- [9] Dinesh Kumar, Pandi Selvam, Palanisamy, "Hybrid Classification Algorithms for Predicting Student Performance", Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS-2021) IEEE Xplore, 2021, pp. 1074-1079.
- [10] Jaikumar M. Patil, Sunil R. Gupta, "Extracting Knowledge in Large Synthetic Datasets Using Educational Data Mining and Machine Learning Models", Soft Computing for Intelligence systems, Springer 2020, pp. 167-175.
- [11] Mustafa Yagc, "Educational data mining: prediction of students' academic performance using machine learning algorithms", 9:11, Springer open, 2022.
- [12] Kamil Yurtkan, Ahmet Adalier, Umut Tekguc, "Student Success Prediction Using Feedforward Neural Networks", Romanian Journal of Information Science and Technology, Volume 26, Number 2, 2023, pp.121-136.
- [13] Noora Shawareb, Ahmed Ewais, Fisnik Dalipi, "Utilizing Data Mining Techniques to Predict Students Performance using Data Log from MOODLE", KSII Transactions on Internet and Information Systems, VOL. 18, NO. 9, 2024, pp. 2564-2588.
- [14] Zara Ersozlu, Sona Taheri, Inge Koch, "A review of machine learning methods used for educational data," Education and Information Technologies, Springer Link, 2024.
- [15] Chitra Jalota, Rashmi Agrawal "Analysis of Educational Data Mining using Classification", International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), IEEE publication, 2019, pp.243-247.