Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 5 Issue 2 (2025)

Prediction of UHV-STEM based educational framework for holistic and sustainable living using Educational Data Mining

Dheeraj Kumar Singh[1],
Narender Kumar[2]
[1,2]Department of Computer Science, Doon University, Dehradun, Uttarakhand, India

ABSTRACT
The rate of digital literacy and computational literacy has undoubtedly risen over the past few decades. But "being literate" rather than "being educated" is still the goal. The focus of every government in the world is undoubtedly on STEM education, and students are also demonstrating positive attitude towards STEM learning. STEM education undoubtedly increases human skill. According to New Education Policies around the world, life skills were essential for the entire development of people. Universal Human Values (UHV) education must now be incorporated into every educational discipline as a course, programme, and subject alongside students' STEM learning disciplines in order to forecast their learning ability or behavior. Undoubtedly, combining the study of science, technology, engineering, and math with UHV (UHV-STEM or UHVSTEM) can have a positive, long-lasting effect on society by removing stress from everyone's life as time goes on. Students with STEM education can behave more predictably by gathering UHV-based Self-Assessment Batteries and using tools and techniques of Educational Data Mining (EDM). This paper will focus on education impact, prediction of course, programme, subject, using k-means algorithm of EDM processes, tools, techniques and benefits of it in diverse classification and clustering of students' performance prediction. The outcome has been taken as focusing on qualitative research with students, teachers through various logical research questions, and proposed UHV-STEM based Student performance assessment model. In this sequence, the fourth sustainable development goal of the United Nations can be satisfied by value-based education, or UHVSTEM education.

Keywords: Educational Data Mining, STEM, Clustering, Universal Human Values, UHV-STEM.

1. Introduction
How we work, communicate, engage in social and educational activities, and have fun is changing as a result of the huge impact that digital technologies are having on economies and society. Today, a wide range of sectors are encouraged to innovate thanks to digital technologies and online education and learning.

The level of digital literacy in the population has a big impact on how well technology can be used for education and innovation. It makes sense that acceptance and usage of digital technology in various sectors of life are strongly correlated with education, skills, and adoption. Creativity can only be fostered with education and skill development.

Massive investments in ICT (Information and Communication Technology) in schools and STEM (Science, Technology, Engineering, and Math) (Xu, W. et al., 2022; Wright & Delgado, 2023; Mumcu, F.,et al., 2023) promotion in schools have not yet resulted in the desired transformation of educational practices and tranquil holistic living, most likely because equally effective strategies for enhancing teachers' professional development, students' personal and professional development, changing pedagogies, and developing appropriate software and courseware have yet to be implemented. Skill

education is at own its place but rather than skill, behavioral aspects are much more important (Romero & Ventura, 2010).

Innovation cannot take place in a vacuum; it requires transparency and links between systems and their environments. This basically applies to education as well. In addition to policies, schools require cooperation from various stakeholder groups to successfully finish the difficult transition process. In recent years, the developing education sector has gained a key role. Instead of being limited to commercial corporate interests selling goods and services to schools, this function is increasingly being regarded as a far larger concern for actual innovation.

The "Vidya Dadati Vinayam" means "Education gives humbleness" or "Knowledge Generates Humility" taught in earlier eras in Indian education system in contrasts with modern events can definitely explore through Artificial Intelligence (AI) and Machine Learning (ML) techniques, but missing of humbleness coming in seen across the world while the literacy rate enhancing across the world. Before 30 years, in general it was the concepts in human being that the teachers, and the doctors considered as God, but in today's world, it is not happening so, or we can say the reduction in values are in scene. Inclusion of UHV in the core of heart of everywhere either in STEM or with NON-STEM learning will help the recall back the same thinking of the students' and of the societies of the world and also will help to reach the world from "one family" to "world united family".

In order to enhance STEM education methods, only items like AI-based robots and 3D printers are marketed in schools, but according to the New Education Policies across the world, enhancing children's values and life skills, including STEM, must be the primary goal. It is possible to effectively alter people's views through UHVSTEM training and investigation employing tools and techniques of educational data mining. Not only marks and behavior can be predicted using Educational Data Mining tools and techniques but also course, programme, and subject predictions are also feasible.

Students can take benefit from educational data mining tools, techniques, and methods in the digital world by concentrating on recent publications of student grades and behavior predictions as well as "Innovative Learning Environments," "Open Education Resources", and "Innovative Learning Strategies for Education and Training." UHV-STEM education, which EDM can assess and better forecast, can lead to a full and sustainable growth of students with their attitude development towards humbleness in self and ultimately lead to harmony in society.

In the end, education based on UHV-STEM can advance Humanity as a whole and contribute to India and the world's peaceful, comprehensive, and sustainable development. Unquestionably, UHV will establish a hierarchy of importance between Right Understanding, Right Relationship, and Right Physical Facility (Gaur et al., 2010) and lead to the understanding of Human Beings and the cohabitation of all four orders of existence material order, plant order, animal order, and human order.

In learning and educating the students, a variety of factors including automatic, adaptive, information gain and delivery, course, programmes, subjects, outcomes, medium, environment, collaboration, complementarities, attitude, motivation, culture, gender norms, religion, sect, caste, cognitive abilities, economics, and society played a significant role. In the current environment, following the pandemic situation of the corona, technology and learning tools are also playing an important role in the learning of students. Artificial intelligence, machine learning, and educational data mining—all of which are byproducts of data mining—have a significant influence on these approaches. Even cutting-edge AI-

based systems like ChatGPT, GEMMA can perform wonderfully in the learning space when combined with Educational Data Mining tools and methodologies.

It is unquestionably true that using digital technology constantly will harm or impair students and kids' memory and writing abilities. However, there is still space being created in the digital world for quick computing, analysis, and task processing. When digitalization improves and students use tools like the AI-based ChatGPT tool, for example, prediction using educational data mining tools and approaches will be important.

The research also solicited responses to the following supporting queries:
(1) What do teachers think about Universal Human Values content for self-exploration?
(2) What do teachers think about the UHV + STEM program?
(3) What are the implications of the UHV + STEM program on lesson planning processes?
(4) What are the implications of the UHV + STEM program on professional development?
(5) What are the implications of the UHV + STEM program on STEM teaching and learning skills?
(6) What do experts think about lesson planning based on the UHV + STEM program?
(7) How should we identify and select most suited course, program, and subject, so that holistic and sustainable education can enhanced a student STEM learning ability? Is this possible by UHV + STEM based course, program, and subjects?

(8) Is Lesson planning can be done for STEM with taking UHV in core of heart of education?
If above questioners can be solved than definitely UHV-STEM education can lead towards sustainability and move towards holistic approach of development.

Are not focusing on the circumstances where students' understanding of values, such as trust, respect, affection, care, guidance, reverence, glory, and love (Gaur et al., 2010), has greater significance than their knowledge of their professional, technical, communicating, teamwork, problem-solving, initiative, enterprise, planning, organizing, and STEM skills.

This paper is organized as follows: the second section will discuss Educational Data Mining, the third section will discuss Literature Review, the fourth section will discuss Aim of Prediction, the fifth section will discuss Proposed Holistic Model for Sustainability, the sixth section will discuss Parameters for Outcome and Implementation, the seventh section will show the Results and Discussion. and the last section concluded the paper.

2. Educational Data Mining

Data mining, commonly referred to as knowledge discovery in databases, used to identify patterns (Agrawal et al., 1993). The term "educational data mining" describes the application of data mining methods for pattern recognition, notably in the educational sector (Baker & Yacef, 2009) (Baker, 2010). A pattern that has been identified in the educational environment and among relevant stakeholders is being used to create and predict future knowledge about student performance, the dropout rate, course and subject selection, and finally programme selection and its accomplishments in the form of awards, rewards, value education, skill development, or enhancement of any or all of these.
Now, everyone who involved in the educational process is curious about the new field of educational data mining. Data mining techniques have been developed in order to automatically uncover concealed

information and trends in educational data. Every year, data is stored in various databases and mined utilizing categorization, grouping, or association procedures (Merceron & Yacef, 2005).

Student modeling, predicting student performance, data visualization, social network analysis, feedback for support management, planning and scheduling, grouping students, detection of undesirable behaviors, and course predictions are some of the application areas for EDM outlined by Romero & Ventura (2010).

Using educational data mining tools and methodologies that can react to the accurate or more appropriate forecast of student performance can improve the learning and teaching abilities of students and teachers; Patterns that significantly alter the forecasting capability of student performance and course selection can be discovered through the use of educational data mining techniques. Vocational Training Programmes can also be researched and student job mapping can be forecasted using Educational Data Mining Techniques.

Data visualization, social network analysis, feedback for support management, planning and scheduling, student grouping, and detection of unintentional conduct / behavior are all topics that can be covered in EDM discussions to help forecast student performance more accurately (Romero & Ventura, 2010).

2.1 Educational Data Mining Process

The process of finding patterns through educational data mining is meticulous and includes processes like educational understanding, data understanding, data preparation, data modeling, evaluation, and deployment (Agrawal et al., 1993; Baker & Yacef, 2009; Baker, 2010; Romero & Ventura, 2010).

Understanding the Educational Context: This involves comprehensively grasping the goals, objectives, and challenges of the educational environment to align the data mining process with specific educational needs.

Understanding the Data: This step entails gaining insight into the data available for analysis, including its sources, format, and quality, to assess its relevance and reliability for the data mining process.

Data Preparation: The data is then cleaned, transformed, and formatted to ensure it is suitable for analysis. This includes handling missing data, removing outliers, and normalizing the data to enhance its consistency and accuracy.

Applying Data Mining Techniques: Next, data mining techniques such as clustering, classification, text mining, decision tree, regression, and association rule mining are applied to the prepared data to uncover patterns and relationships.

Model Evaluation: Evaluation metrics such as accuracy, precision, recall, and F1-score are used to assess the performance of the models and determine their effectiveness in addressing the educational goals.

Deployment of Findings: Finally, the insights and findings from the data mining process are implemented in the educational setting to inform decision-making and improve educational outcomes.

2.2 Techniques of Educational Data Mining

2.2.1 Text Mining

Singular value decomposition (SVD) is a method for dissecting a matrix into a number of approximations. Singular value decomposition (SVD) is a technique for processing data sets with millions of rows and dozens of attributes. For text mining, this technique performs reasonably well (Liu et al., 2005). Using a variety of tools, APIs, and software, text mining locates, analyses, and categorizes textual material. Using tools, text mining expresses representational linkages and extracts data from the syntax, word meaning, and parts of speech of texts (Slater et al., 2017).

2.2.2 Decision Tree

A decision tree is a non-cyclic, tree-structured flowchart for presenting information and formulating predictions. A decision tree is an intelligent tool for supporting decision-making when developing derivation rules. Internal and exterior nodes are both always leaf nodes and have the ability to forecast. Non-leaf nodes are always internal nodes. Due to the graphical character of decision trees (Merceron & Yacef, 2005; Asif et al., 2017; Fernandes et al., 2019), they have evolved into a mechanism for showing entities as nodes, such as root nodes, internal nodes, and leaf nodes.

2.2.3 Classification Analysis

The performance of students can be predicted more precisely using techniques for classifying data. The qualities of the students enrolled in a particular course have an impact on how effective a classification system is. According (Abu Tair & El-Halees, 2012), one of the best methods for classification is the decision tree. In essence, classification organizes data for effective and efficient use according to time-space trade-offs, techniques, and methodologies used in it (Durairaj & Vijitha, 2014). A more precise forecast of student performance can be made using the Naive Bayes classification method, where one particular feature is always present independently of other features (Devasia et al., 2016).

2.2.4 Clustering Analysis

Clustering is an unsupervised learning technique that is useful for educational data mining and has a wide range of other applications, including scientific data exploration, text mining, spatial data analysis, medical diagnostics, computational biology, web analysis, CRM, marketing research, and many others (Durairaj & Vijitha, 2014). The information at hand can be used to study patterns and trends utilizing this data mining technique. A collection of practical features and realistic patterns can be produced for instructional data mining. Implementing and developing statistical approaches or algorithms employing clustering methodology can help students perform better across disciplines (Pena-Ayala, 2014).

3. Literature Review

Prediction and classification techniques are related to one another. Classification forecasts the labels for each class, but prediction reveals continuously valued functions. There are numerous web mining techniques for educational systems, but almost all of them fall into one of the following categories: text mining, association rule mining, sequential pattern mining, clustering, classification, and outlier identification (Romero & Ventura, 2007).

EDM makes it feasible to improve domain knowledge models, learning software system pedagogical support, course support, and student performance models. Data from keystrokes, answers, sessions, students, classes, and schools, as well as from lab tests and design study, can all be taken into account

in scientific learning research. Time, sequence, and context are crucial considerations when analyzing student instruction (Baker, 2010).

When the values of a target feature are unknown, machine learning techniques referred to as "unsupervised model induction" infer models from training data. Unsupervised approaches use a bottom-up strategy to find patterns and structures in the input space without the use of explicit goal categories or tagged samples. Clustering is a popular method for identifying groups of examples in a training set that are "similar" in some way, and collaborative learning may be an effective learning strategy (Scheuer & McLaren, 2012).

Some of the most common EDM problems are the incremental nature of educational data, a lack of data interoperability, data ambiguity, and a relationship between student and teacher research knowledge (Jindal & Borah, 2013). To determine a test's difficulty using regression and classification predictive algorithms, respectively, EDM techniques can be employed in automated test administration and preparation in programming (Ivanevi et al., 2014).

Two techniques can be used to categorize the data in semantic constructs: one employs bottom-up "unsupervised learning" techniques like clustering and the other uses top-down "supervised" prediction techniques like classification and regression used in educational data mining. Neuro-fuzzy classification is the most popular classifier technique in educational data mining. (Buniyamin et al., 2015) used data from a Malaysian university to predict and categories' students' academic progress. EDM is a new approach in data mining and knowledge discovery in databases (KDD) that focuses on discovering useful patterns and pertinent information from educational information systems (Saa & others, 2016).

By applying sophisticated machine learning algorithms, the K-means clustering technique can also be used to observe the pattern of educational data (Fernández & Luján-Mora, 2017). The Multilevel Perceptron (MLP) classifier performed marginally better results than the Nave Bayes (NB), Nave Bayes Updateable (NBU), Decision Table (DT), Decision Stump (DS), Decision Tree (J48), Simple Logistic (SL), RandomTree (RT), Random Forest (RF), and REPtree (RepT) classification algorithms (Zaffar et al., 2017).

Some of the most popular algorithms for educational data mining nowadays are decision trees, neural networks, and Bayesian networks (Fernandes et al., 2019). Tutoring staff without a background in data science can examine and forecast student performance in terms of observable scores and finishing their studies with the use of a web-based software application for student profiling. This tool can be used for drop-out prediction, course prediction, exploratory analysis, grouping, and classification (Prada et al., 2020).

Ensemble learning methods make predictions in educational data mining that are more accurate than those from individual classifiers. Base classifiers including Random Tree, Decision Tree-j48, KNN, and Naive Bayes, as well as filtering strategies like oversampling and under sampling, can be analyzed to forecast students' performance and develop an accurate prediction educational model (Ashraf et al., 2020). Scientific reasoning, which includes developing hypotheses, running tests, and analyzing data, can be used to find variables when employing Educational Data Mining techniques (Arnold et al., 2021).

Process mining and expert feature engineering can be used to anticipate for the students if they are using their time successfully or not based on educational data mining approaches and machine learning methodologies (Levin & others, 2021). By employing clustering and temporal data mining techniques like process mining and sequence mining, it may be feasible to profile the implementation of Self-Regulated Learning (SRL) when using EDM approaches. Clustering is now the most often used implementable method for SRL profiling, according to studies. Real-time learner profiling may be made possible by integrating EDM techniques with learning management systems (Araka et al., 2022).

(Xu, W. et al., 2022) with the emergence of AI technology in STEM education, the issue of how to incorporate various AI strategies into the current STEM educational system has come up. The scientific literature on AI and STEM was evaluated as part of this study, and recommendations for the theoretical, technological, and educational implications of AI application in STEM education were also made. In general, there is a great need for additional research into how technology affects educational systems and how AI technology may help to advance STEM education.

(Wright & Delgado, 2023) study about LGBTQ (Lesbian, Gay, Bisexual, Transgender, Queer) and found that the problems are being created across the world not even in USA in the educational institutions or universities and students and teacher culture are not in safe hand, and proposed a model of education contains framework for gender sexual diversity-inclusive STEM (GSD-inclusive STEM) education, so the fear can be eliminated among the students and they can feel safe in education environment itself, GSD they proposed as a language for educational environment with STEM learning.

(Mumcu, F.,et al., 2023) studied and focus onto the computational thinking with STEM mergence using qualitative and quantitative approach through Likert Scale / parameter using own framed different reasonable research questions and ultimately guided for making lesson plan in integrated STEM environment.

(Gupta, M. K., & Tiwari, M., 2022) study on male prisoners clearly shows that crime of every kind may be reduced minimum upto 20.25 percent when every human being passes through the content of Universal Human values (UHV) and done a quantitative analysis in the jail with mail prisoners.

4. Aim of Prediction

In this section, ongoing research and prediction earlier the skill-based education has been identified and prioritized more, due to that sustainability has become crucial in the organizations. Before prediction of Education through Educational Data Mining tools and techniques, it is necessary to discuss here the definition of “Education” and “Sanskar”. As per study research question generated today:

RQ: Is Role of Education for Holistic development and transformation to Human Consciousness?
The role of “Education-Sanskar” is to enable this transformation by way of ensuring the development of the competence to live with human consciousness and definite human conduct
For this, it has to ensure

1. Right understanding in every child
2. The capacity to live in relationship with the other human being
3. The capacity to identify the need of physical facility,

the skills and practice for sustainable production of more than what is required – leading to the feeling of prosperity (Gaur et al., 2010)

As per oxford definition "School" is "a place where children go to be educated" and "Education" is "a process of teaching, training and learning, especially in schools, colleges or universities, to improve knowledge and develop skills"

So, as per the oxford dictionary "Education" definition leads only towards the skill development and while we take the reference of Gaur et al., 2010 the "role of education" leads towards the holistic development of students as well as all Human being. So as per study oxford definition for education leads only towards the body centric development, while at another end Gaur et al., 2010 definition for education leads towards the self-centric development, so lead towards the harmony in Human being. Educational Data Mining can be useful for exploring and predicting such sustainable and holistic education based on Gaur et al., 2010 and for prediction of behavior of students. Some of the education strategies are discussed below for prediction of better education in the world. Here values centric approach for sustainable education cum development also has been discussed.

4.1 Skill Based Education

Still, the education system whatever being followed in the world is related to maximum skill based education, due to this system of education students only predict the courses according to seeing the research questions such as:
RQ1: How economical the course, subject, and program are?
RQ2: How much economics they will get after completion of course, subjects and Program?
RQ3: How quick relocation is possible in that field of learning for economical growth?
Even this skill-based learning sometimes leads towards the failure to sustain the students or learner (employee), even while this student or learner knows that moving him can lead towards the product failure, sometimes the organization failures. So, if see education scenario prediction of course, subject, and program by student is economic, skill centric nowadays.

4.2 Values Based Education

(Gaur et al., 2010) said that all the universal human values parameters are universal, time invariant, rational, verifiable, and leading to harmony. Change in geographical locations has no impact on parameters and learning of Universal Human Values. So, if the students have to be educated in such a way, when core of learning in students is based on Universal Human Values than such type of students will work with more stability in the organizations, behave well, and will generate the harmonious environment for the society. Since Values directly related to the behavior of the students, so by applying or generating different psychological assessment batteries, the behavior of students also can be predicted. Since psychological assessment batteries will generate the parameters which will be categorical or numeric in nature so classification, clustering, and related techniques also can be applied for identifying the behavior of students, and right understanding can be ensured. As well, right utilization of acquired skill can be done by understanding Universal Human Values.

4.3 Prediction of Course, Program, Subject, Marks, and Behavior

According to a study, learning environments like schools, colleges, universities, and educational boards are increasingly focused on student-centered, body-centric, and skill-centric learning approaches in India and around the world. They place more emphasis on the promotion of STEM

fields, whereas character development programmes, courses, and subjects are becoming less prevalent in the current educational system. These organizations place the utmost emphasis on students' ability to achieve ranks, a high percentage of passing grades, and educational organizations also focus on this in news for more admissions globally. Character development is definitely absent, and eventually, human behavior is deficient everywhere in the world.

Additionally, it has been discovered that many businesses around the world seek out students who can work well in a team, last a long time in an organization, show commitment to the work and family culture there and in society, distinguish sustainably between the cost and value of a product as well as relationships, and live with definite human conduct. All of these things are made possible through good character development, and by setting the priority of right understanding, relationship, and understanding physical facility (Gaur et al., 2010), ultimately will lead this towards the good behavior, which will also consider more parameter in every self for behavioral performance of the students as well as teachers.

Without a doubt, educational organizations must concentrate on the STEM learning education based on universal human values throughout the world as well as the teacher-centric, value-centric, and self-centric learning of students throughout the world, including India, which can ultimately result in good human behavior. With the development of this socioeconomic society of the world, educational data mining methods with AI and machine learning, tools, and techniques will not only be useful for predicting student grades and rankings but also useful for predicting student behavior and the inclusion of UHV as a course, programme, and subject in every discipline of education. Inclusion of UHV will lead towards the transformation of character building in every human being, and STEM exploration will lead them to make more economics, which combined will define holistic and sustainable growth versus development in the society.
The study must concentrate on at least six factors:
(1) UHV-STEM education in traditional regular learning, distance learning, and online learning,
(2) Value acquisition, Lesson Study (LS),
(3) Lesson preparation procedures,
(4) Lesson planning obstacles, and challenges
(5) Integration, evaluation, and assessment techniques
(6) Ideas, strategies, tactics, and procedures.

4.4 Proposed Comparison Table between STEM and UHV-STEM Education with role of AI / ML and EDM Approaches

(Table 1) shows by implementing UHV based STEM education system in institutions / organizations such as schools / colleges / Universities the prediction of sustainable and holistic learning can be possible, and finally this type of learning approach will create harmony towards the humanity and then development occur across the world will be sustainable. Through Artificial Intelligence (AI) / Machine Learning (ML) exploration of such education will be faster and by applying the EDM tools, techniques, and approaches the prediction of marks, behavior, course, and jobs can be done.

(Table 1) is the outcome based on the exploration of Universal Human values by (Gaur et al., 2010). A skill guided by value can lead harmony in human being. It is possible by exploration of content of Universal Human Values (UHV), and can satisfy the fourth sustainable development goal of the United Nations. Since, UHV is the value-based education that is one of the reasons to acquire it by All India Council for Technical Education (AICTE), University Grant Commission (UGC), and ultimately

by Ministry of Education, government of India. It shows, ultimately UHV education will make students to live with definite human conduct.

Table 1: Comparison between STEM Learning and UHV-STEM based learning education System for Sustainability

Learning Approaches	**STEM LEARNING**	**UHV-STEM LEARNING**		
Focus	Acquiring More Skill to earn more money – No True Socio-economic Growth Experiencing Speciality in individual	Acquiring Life –Living Skills and value for money - Actual True Socio-economic Growth Actualization of Excellency for all Human being	Artificial Intelleigence (AI) / Machine Learning (ML) may be applied for exploration of UHV-STEM fastly across the world in institutions / Organizations including schools / colleges / Universities For defining new course / Programme / Subjects	Using Educational Data Mining (EDM) tools and techniques prediction strategy can be developed for Marks as well as for behavior and Job Mapping can be done for holistic and sustainable growth
Approach	Body Centric Approach Limited upto- to learn (skill) -> to do-> to taught	Self-Centric Approach Exploration upto- to learn (skill) -> to do-> to taught and to understand (values) -> to live accordingly-> to make understandable by living rightly		
Behavioral Impact	Nothing	Leading to Harmony in Life		
Ability Enhancement In the form of	Enhancing Competition with others	Enhancing Competency and complementarity with all		
Student's Behavioral Responsibility	- According to Family it is the responsibility of Teachers - According to Teachers it is the responsibility of Student's Families	Complete Responsibility of Right Education System		
Conduct	No talk about conduct	Definite Human Conduct		
Concept	Current Education System	Indian Knowledge System		

4.5 Predicting behavior through Sentimental Analysis through EDM

By applying UHV-STEM based model for educational environment can give a new dimension for holistic living of human being. As per (Babu, G. at al., 2022) understanding co-existence and live accordingly will unfold the holistic world view for complete Humanity. Human always lives at four levels identified by Gaur et al., 2010 that are living at the level of self, family, society, Nature and Existence. At the level of self, Human is the co-existence of self and body and the need of self and body are both different and since need of self and body are different, so sentimental analysis can be applied on to the different needs of self and body and finally the classification and clustering approaches of Educational Data Mining can be applied to predict the student's behavior. Parameter analysis is the most complicated part when applying Educational Data Mining classification and clustering approaches for behavior prediction.

4.6 Performance Metrics

In the methodology, final goal become possible to hit through taken confusion matrix and parameter associated with the matrix named as Accuracy, Precision, Recall, F1-Score, and Specificity (Xia, 2020), which are given as:

True Positive (TP): The number of correct predictions that an instance is positive.
True Negative (TN): The number of correct predictions that an instance is negative.
False Positive (FP): Incorrectly predicted as positive while it is negative.
False Negative (FN): Incorrectly predicted as negative while it is positive.

Accuracy = (TP + TN) / (TP + TN + FP + FN) (Formula 1)
Precision = TP / (TP + FP) (Formula 2)
Recall (Sensitivity) = TP / (TP + FN) (Formula 3)
F1 Score = 2 * (Precision * Recall) / (Precision + Recall) (Formula 4)
Specificity = TN / (TN + FP) (Formula 5)

5. Proposed Holistic Model for Sustainability

When Universal Human Values establishes in the core of human being or students and then learning of Science, Technology, Mathematics, and Engineering , overall known as UHV based STEM Learning (UHV-STEM learning) can lead to sustainable and holistic education, and can make the effort of students towards the harmonious living of every human being onto the earth, and finally a student, who govern by such UHV based education will lead to the sustainable and holistic development of the origination and the world (Figure 1). This type of education system based on Universal Human values in the core of heart of Human-being can always will generate the living of Human being with feeling of right understanding and will generate the harmony in self, family, society, and nature / existence. So, proposed model Figure 1 may be fit for implementation across the world.

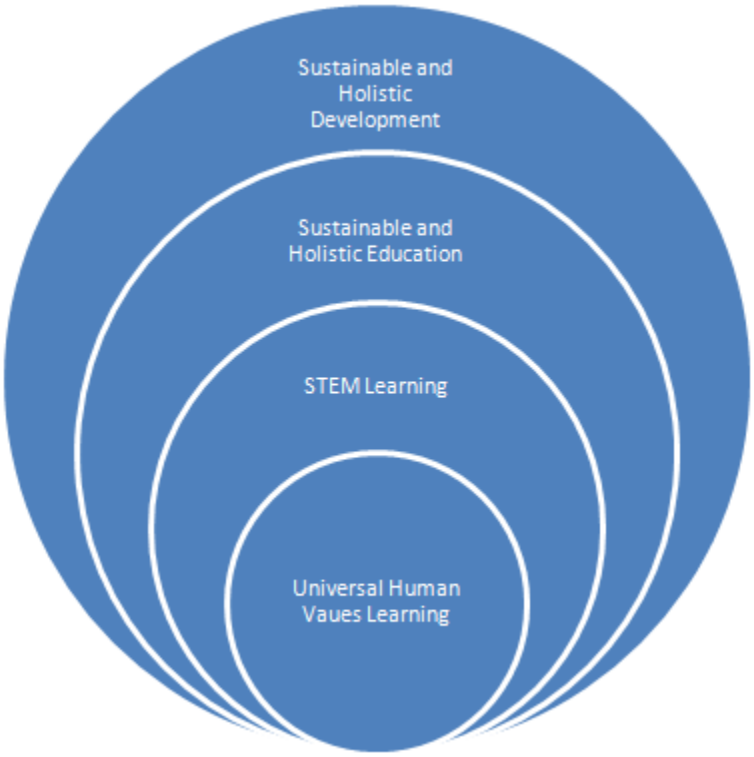

Figure 1: UHV-STEM based learning education model for Sustainability

6. Parameters for Outcome and Implementation

6.1 UHV Language

The below study in step 7 with the students, and also the results of the (Gupta, M. K., & Tiwari, M., 2022) shows that as per our study implementation of the UHV content as a UHV language with proposed UHV-STEM framework (Figure 1) in the educational environment definitely can lead towards the harmony in every self, family, society, nature, and existence. The framework proposed in Figure 1 will lead towards the holistic and sustainable education approach and UHV-STEM based this learning will lead towards the holistic and sustainable development.

6.2 Proposed UHV-STEM based Student Performance Assessment Model

Student's performance not only dependent on self but also have combined effect of self as well as the community Figure 2, where ultimate goal of the human being is to live with continuous happiness "Sukhi hokar jeena" means "to live happy" (Gaur et al., 2010), and is the one of the major parameters for student as well as every human life and part of the ultimate performance. By doing and applying Educational Data Mining tools, techniques, and methods such as text mining, decision tree mining, classification, and clustering prediction of student behavior or sentimental analysis, as well student's educational grade prediction is possible, student's team work towards sustainability can be measure through taking his time series analysis during student's data set available during their education in Kinder Garden, Primary (K1 to K5), Junior (K6 to K8), Secondary (K9 to K12), Higher Education (Graduation and Post Graduation) (K13 to K15 [,K16])[,[K16 TO K17]] and finally Doctoral programs done by the students. So, studying of Self feature of the students such as their understanding towards Universal Human values (UHV), Science, Technology, Engineering, and Mathematics (STEM), Self and Body (Gaur et al., 2010), co-existence in all units presents in existence, and

Community feature such as family, society, school, college, and university and inherent interconnectedness (Babu, G. at al., 2022) among all, will lead towards socio-economic ecosystem, and will lead harmony in human being / student at every phase of life, and ultimately will lead the sustainable education cum development to the globe.

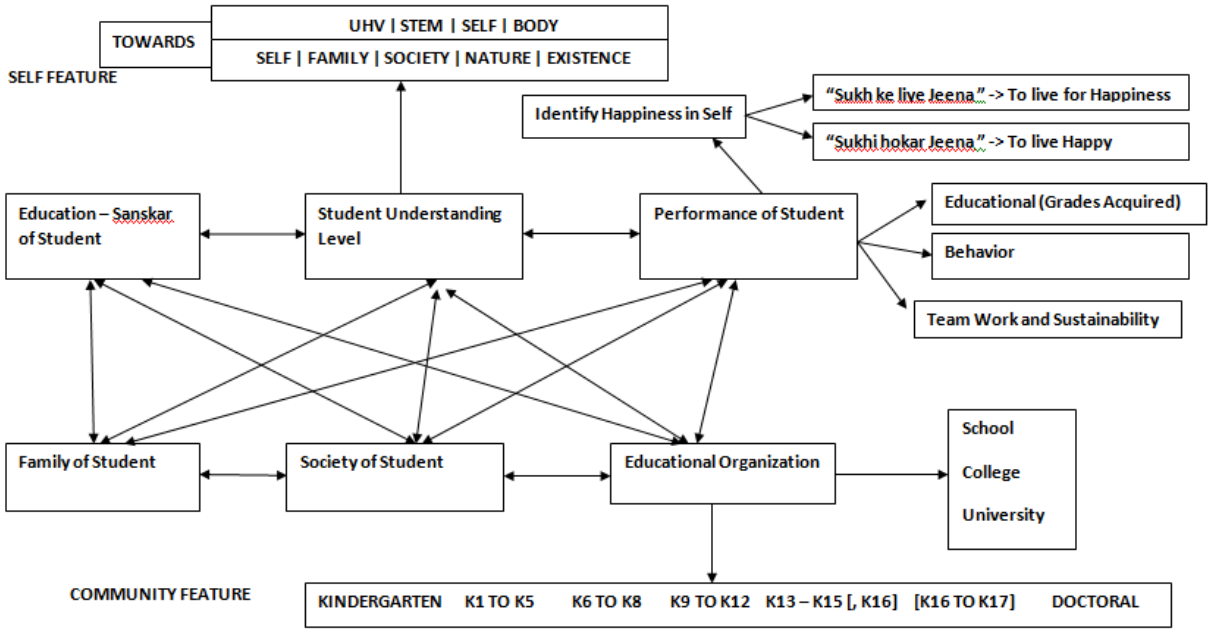

Figure 2: UHV-STEM based Student Performance Assessment Model

6.3 Cultural and holistic view of UHVSTEM model

As the inhuman behavior of the human being is increasing across the world among the most literate people, and it has been many time, and all the newspapers and media is basically the witness for this inhuman behavior, and can be measured through the news that education institutions are increasing, family fueds, hospitals, number of courts are increasing across the world, while as the education is improving it must be lesser again and again and humanity among society , co-relation, and collaborative work environment, decentralization of physical facility have to be enhanced. So, the Human being can live with basic aspiration of continuous happiness and prosperity, but it is not being done, so in this time this research discussed in Figure 2, will move human being to live with definite human conduct. By following proposed UHVSTEM model of education will move the students towards understanding the holistic and sustainable socio-technological development, which will not only focus on socio-economic development but also will focus on socio-ecological development. Exploration of UHV language will guide every students / human being towards understanding of the co-existence present in all four orders, and making sustainable technological development of the world. UHVSTEM implementation in education will develop the holistic and sustainable culture across the world.

6.4 Machine Learning Encoding Scheme Used for prediction

One-hot encoding transforms category input into a numerical format, which is useful in machine learning. Likert scales and other multi-level categorical variables are especially effectively managed with this tool. On a 5-point Likert scale, from "Strongly Agree" to "Strongly Disagree," each category is represented as a binary vector. With 1 denoting its presence and 0 denoting its absence, each

category turns into a binary feature. One-hot encoding makes sure that there is no implied ordinal relationship between the categories in the numerical representation, protecting the integrity of the original data. Five binary columns are constructed, one for each category in a Likert scale dataset. One-hot encoding allows machine learning algorithms to handle each category independently without assuming a predefined hierarchy or relationship between them. After the encoding process is completed, various machine learning algorithms, for classification such as Random Forest, Decision Trees, Support Vector Machines (SVM), K-Nearest Neighbours (KNN), and Logistic Regression, can be used depending on the goals in order to produce a confusion matrix and correlation matrix. These can then be used to calculate the accuracy, recall, precision, and F1 score (Cerda et al, 2018). After applying One-hot encoding clustering techniques also can be used for prediction and analysis of Likert scale data such as k-means clustering, hierarchical clustering, Gaussian Mixture Models (GMM). Finally, the scatter plot, heatmaps, ROC curve can be generated for the clustered data which will give the perfect visualization for the prediction and demand of the UHV language for the holistic and sustainable behavior of students towards the living of life with harmony.

6.4.1 k-means clustering

As per (Raval & Jani, 2016), Clustering techniques play a vital role in data analysis, and among them , k-means stands out as one of the oldest and most widely used methods. It involves two main steps: deriving initial centroids and assigning data points to their nearest clusters. In the basic K-means algo rithm, the choice of initial centroids significantly impacts the resulting clusters. Initially, K data elem ents are selected as centroids, and then the distances of all data elements are computed using the Eucl idean distance formula. Data elements closer to centroids are grouped into the corresponding clusters . This process iterates until no further changes occur in the clusters.

K-means clustering is a popular technique that requires the user to specify parameters such as the nu mber of clusters (k), cluster initialization method, and cluster metric. Initially, the algorithm defines c lusters by creating subsets or groups of points closest to the centroids within the dataset, which are th en labeled as clusters. Next, it calculates the mean value for each cluster to determine a new centroid and reassigns data points to this new centroid. This iterative process continues until the centroids stab ilize. The basic algorithm for traditional K-means is as follows:

Input: D = {d1, d2, d3, ...dn} // set of n numbers of data points K // The number of desire Clusters
Output: A set of k clusters

1. *Select k points as initial centroids.*
2. *Repeat*
3. *by assigning each data point to the closest centroid among K clusters.*
4. *Until the centroid remains unchanged, recalculate the centroid for every cluster.*

Important Equations

The calculation of two elements' similarity will be based on their distance measure, which will also influence the clusters' shape.

The Euclidean distance is referred by:

$$dist \square \sqrt{\sum^{n}(p_k - q_k)^2}$$

Where n is the number of dimensions (attributes) and p_k and q_k are, respectively, the kth attributes (components) or data objects p and q.
The Flow char-based algorithm is as depicted in Figure 3.

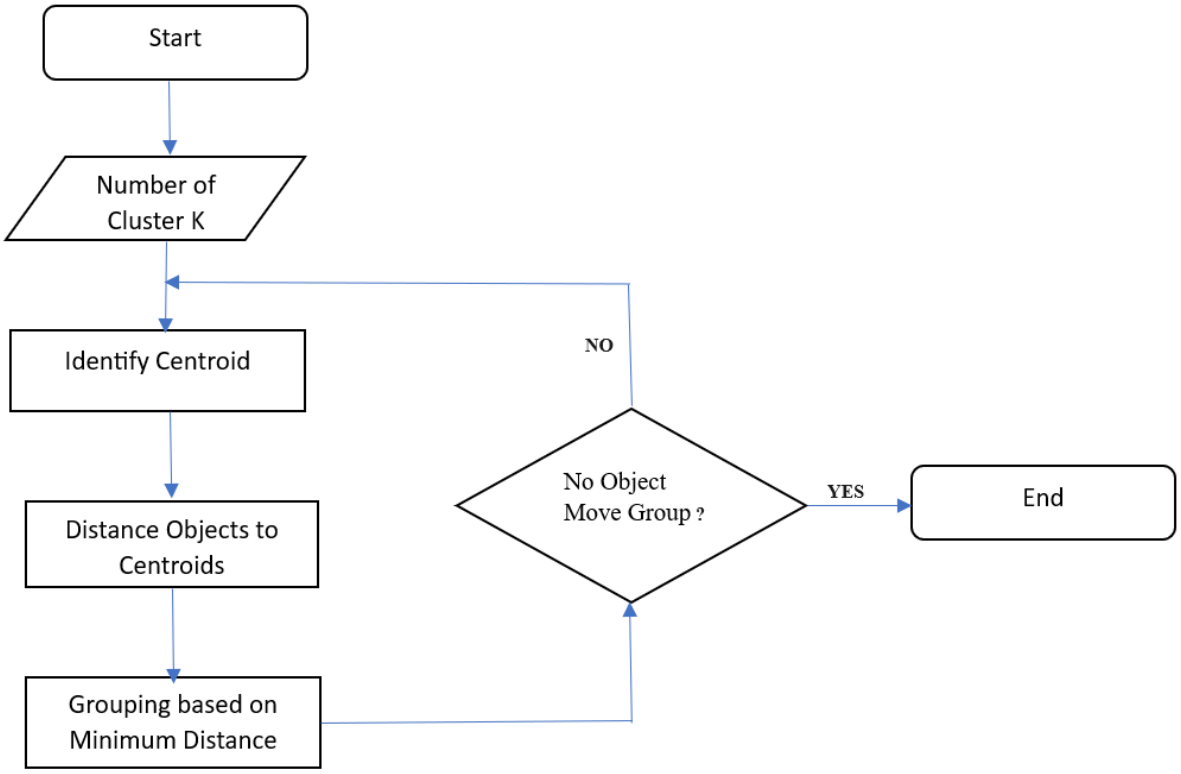

Figure 3: Basic Flow chart for K-Means Algorithm

however, the algorithm has its own pros and cons, which is as follows;

PROs:

1. This clustering algorithm is comparatively faster.
2. It works fast with the large and small data set both and have the time complexity is O(nkl) where n is the number of patterns, k is the number of clusters and l is the number of the iterations.
3. Because it is based on Euclidian distance, it performs well with numerical values that have interesting geometric and statistical meaning.

CONs:

1. varied values for K will result in varied numbers of clusters since the initial assumption of value for K is crucial, but there is no sufficient description on assuming value for K.
2. The first cluster centroids are crucial, but if the centroid is too far from the data's cluster center, endless iterations may occur, which may occasionally result in incorrect clustering.

7. Results and Discussion

In this section, we describe the results obtained in some specific phases of the Research Questions asked. For seeing effective impact of UHV on students and teachers 10 research questions has been asked, where all students were from the science stream and teachers / professors were from art and science stream both. A five-point Likert scale has been followed during assessment taken from students, while filling the assessment from teachers / professors for effective implementation responses have to be collected on binary parameter yes / no. The research questions asked were as:
RQ1. Is it be sustainable and holistic education when incorporation of UHV done with current education?

RQ2. Is UHV based education will satisfy the holistic and sustainable living?
RQ3. Is UHV based education will lead students towards definite human conduct?
RQ4. Is UHV based education will lead students towards sustainable and holistic development?
RQ5. Is UHV based education will lead Harmony in Human being?
RQ6. IS UHV incorporation with current education will lead students towards Happiness?
RQ7. IS UHV important for all students / teachers / society for holistic living?
RQ8. IS only STEM (Science, Technology, Engineering, and Mathematics) education will lead Harmony in Human being?
RQ9. Is addition of UHV education with STEM education will lead Harmony in Human being?
RQ10. Is delivery of UHV "Education- Sanskar" is necessary for all students?

7.1 UHV content Outcome tested on Students

After giving the session on UHV to the students in their regular class for 18 hours, the qualitative assessment being recorded using assistance of google form, and finally the following outcome were came as presented in diagrams below as per research questions asked, in this study all students (195 students) were from graduation first year and from the Science and Arts stream (physics, chemistry, mathematics, geography, geology, media and communication, biological science, and design), and total number of 137 students gave the response and fill the assessment form and outcome of the study for the above said research questions, in this study 90.5% students were from age group 16-19 years of age and rest 9.5% students were from age group 20-23 years of age (Figure 4), and in the study 71.5% female participated and 28.5% male participated (Figure 5). Count of Age graph, and Gender ratio in the study has been generated by using the google analytics.

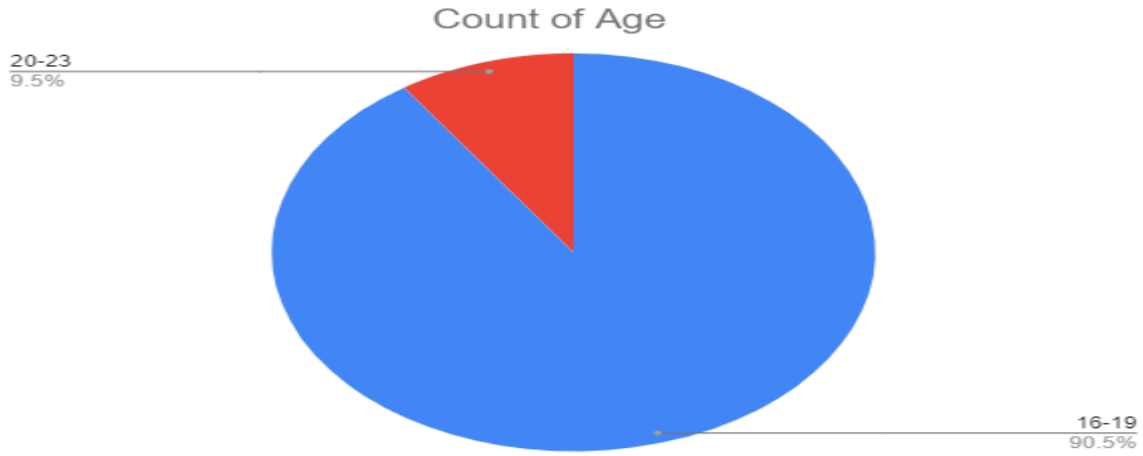

Figure 4: Pie Chart for participant Age

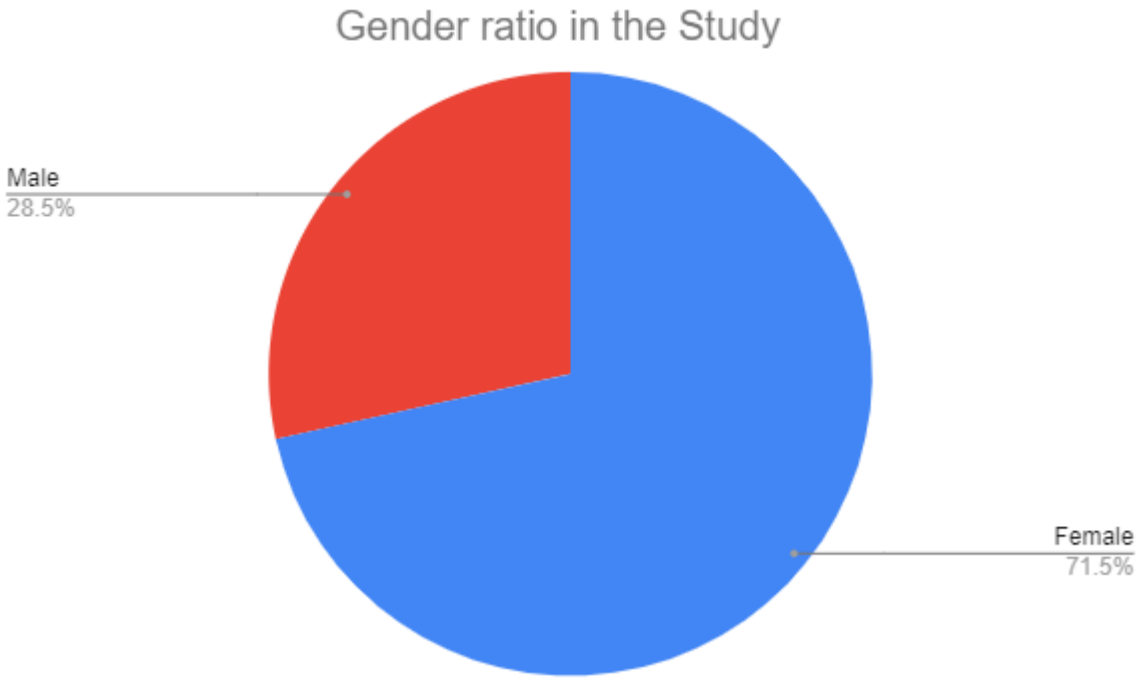

Figure 5: Pie Chart for Gender ratio in the Study

A further stacked bar chart (Figure 6) has been drawn where it is clearly shown that corresponding to all research questions asked to the students the percentage of students 'Agree' towards UHV education is very much high, only the research question RQ8 is there which is based on only STEM based holistic living, where STEM is not possible to lead Harmony in Human being. So, students who attended the sessions on Universal Human Values education identified that this education should be in education system.

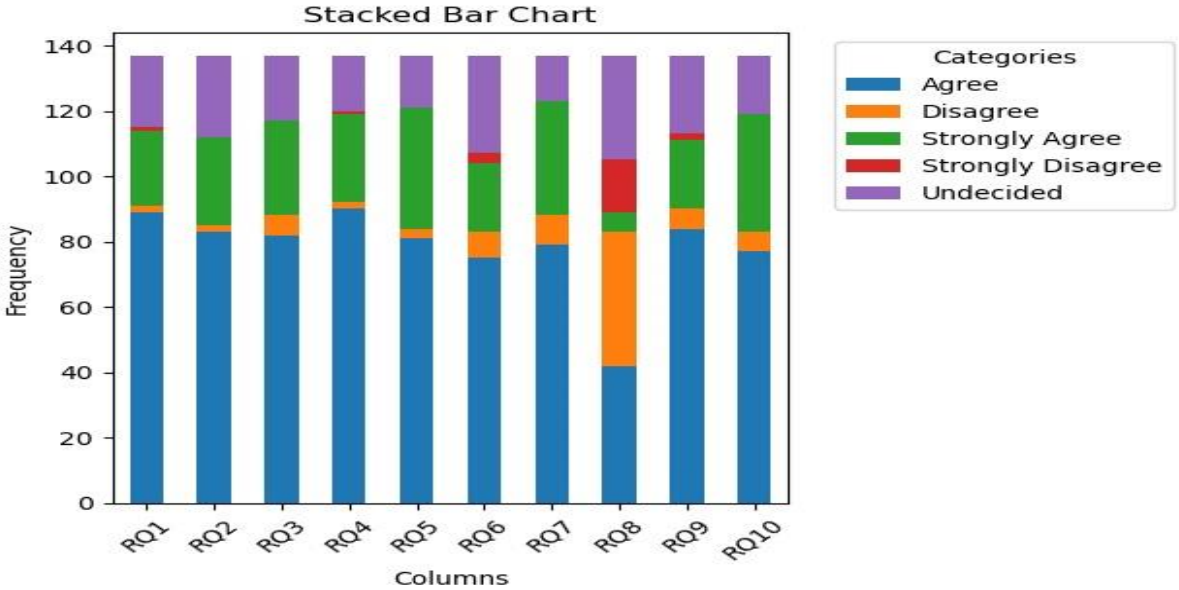

Figure 6: Stacked Bar Chart for showing UHV performance reference to all RQs

7.2 Data Visualization through clustering of Machine Learning

In the above context, research has been done on 137 student's data and k-means (2-means) (Fernández & Luján-Mora, 2017) unsupervised machine learning algorithm has been applied and data has been converted into the 2 separate cluster using the machine learning with python the results were analyzed.

The confusion matrix has been calculated for analyzing that UHV have to be part of Curriculum as a Course, Subject, and Program, this can also be seen by the applying the unsupervised machine learning k-means (2-means) algorithm using the python coding. The results with Graphical Visualization of the confusion matrix identified are shown in Figure 7.

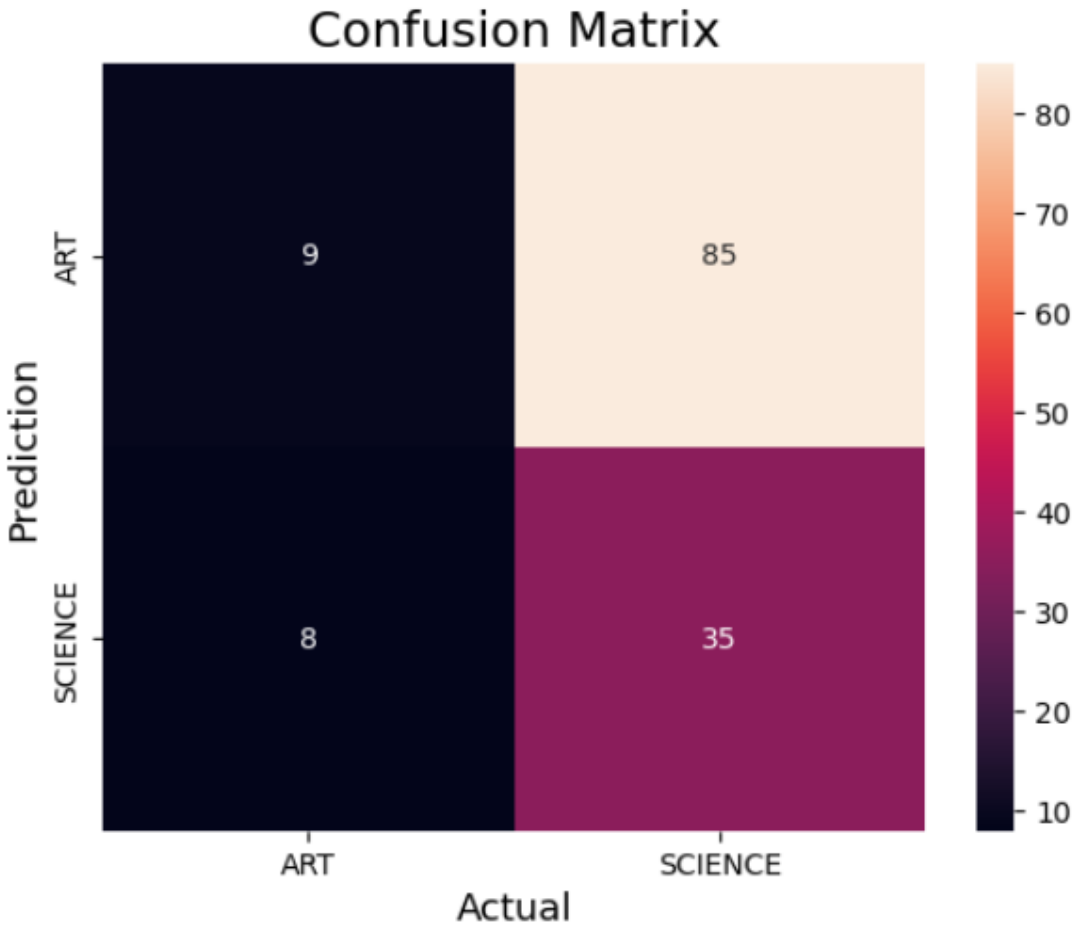

Figure 7: Confusion Matrix for students' adoptions for UHV

Where 85 (approximately 71% out of all 120 Science Students) students of science is saying that UH V have to be the part of education for all and 8 (approximately 47% out of all 17 Arts Students) stude nts are also saying that UHV should be part of education while at another side only 35(approximately 29% out of all 120 Science Students) students of science and 9(approximately 53% out of all 17 Arts Students) students of arts are saying that UHV is not necessary in education. So, it may be possible t hat Science students in general not read subjects of Philosophy in your studies so science students' in clination towards UHV is very much high, while at another end Arts students in general read subjects

of Philosophy so their inclination towards choosing UHV is lesser than that of Science Students. Also, the results of cluster centers indicate that Science Students inclination towards UHV is very much high.

By applying K-Means, the confusion matrix derived some results corresponding to the selection of UHV as a course, programme, and subjects by the students, where it is identified that the accuracy of adoption was 77.5%, Precision was 90.42, Recall value was 70.83%, F1-Score was 79.44%, Specificity was 47.06%, Type – 1 error identified 52.94%, and Type – 2 error identified was equivalent to 29.16% (Figure 8).

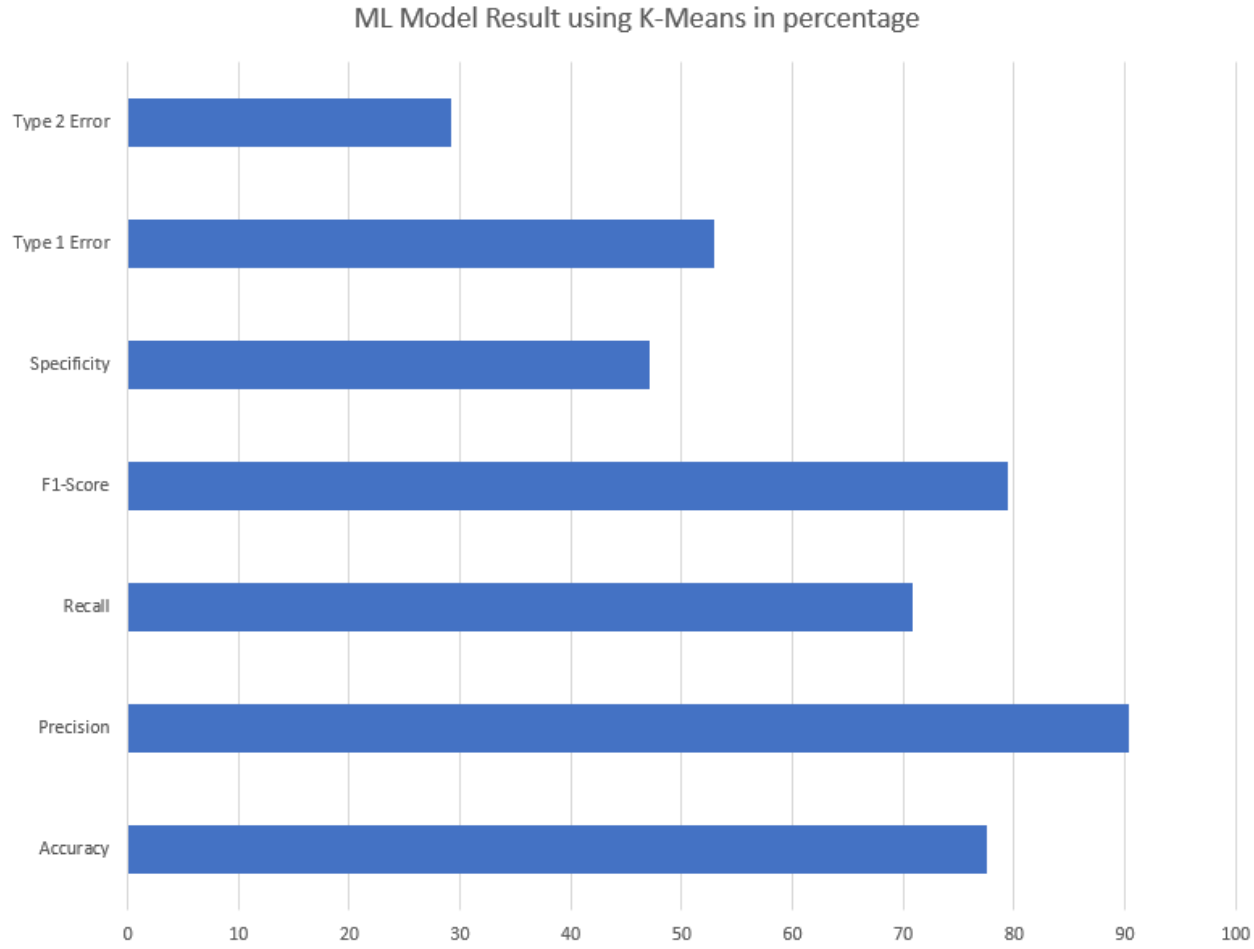

Figure 8: ML Model outcome using K-Means Feature Selection

The further again the Bar chart drawn on-behalf of confusion matrix where the comparison has been shown between the art and science students (Figure 9) which gives clear understanding that adoption rate of science students for UHV as a subject matter is very much high, while over the art students either read somehow the contents related to it so adoption rate of UHV by arts students is lower than that of science students. Ultimately both stream students want that UHV have to be the part of education system.

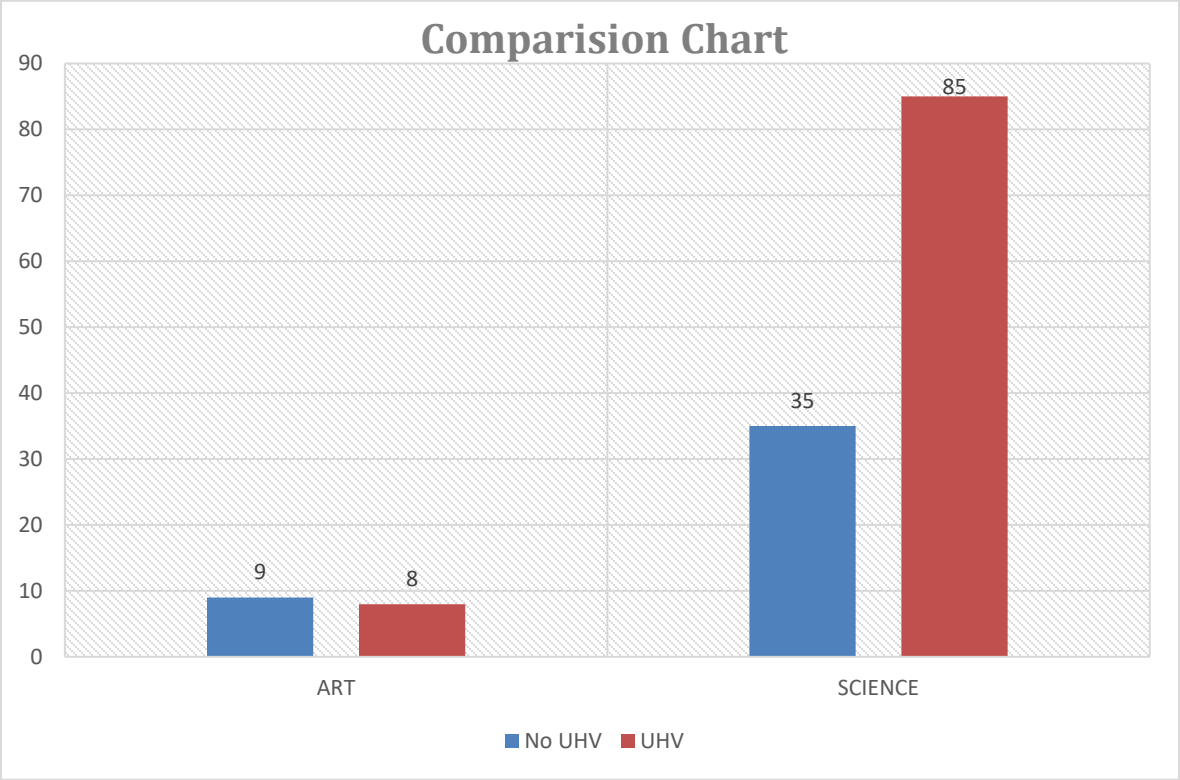

Figure 9: Comparision Bar chart for the ART and Science students proposing UHV

A receiver operating characteristic curve, or ROC curve, is a graphical plot that used for visualize the performance of a binary classifier model (Adekitan & Noma-Osaghae, 2019; Yağcı, M.,2022). The ROC curve (Figure 10) also has been designed for the same adoption rate of UHV by the students an d it is being measured that the True Positive Rate is very much high corresponding to the False Positi ve Rate. So, it clearly says that UHV must be the part of education system, since after the one point o f time sensitivity corresponding UHV is increasing very fast. Even applying the confusion metrics fo rmulas, sensitivity towards UHV inclination is more than 90%.

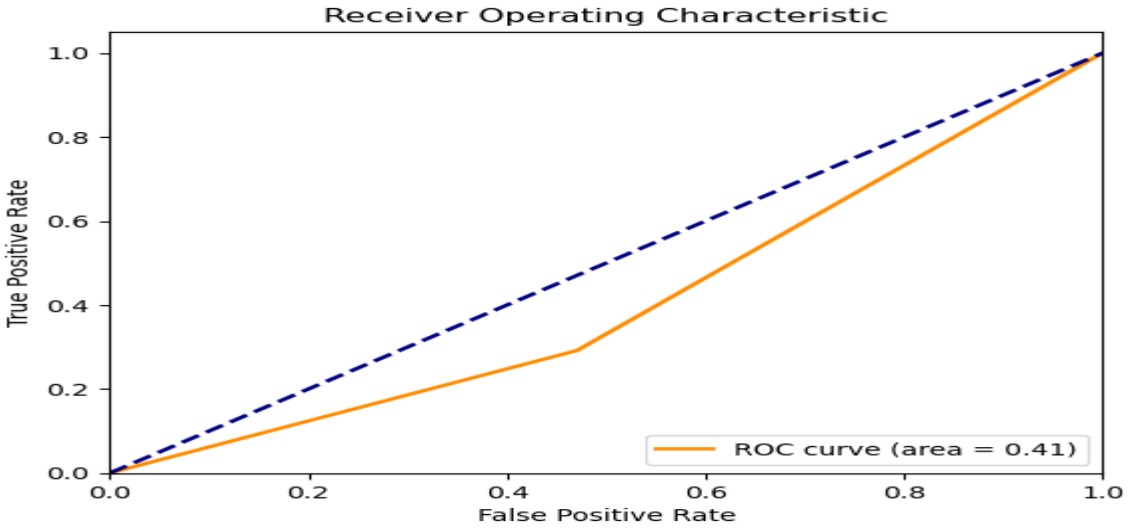

Figure 10: ROC curve for the ART and Science students proposing UHV Growth

The outcome given by teachers / professors clearly shows it will be beneficial to implement UHV with STEM for better development of educational environment and will lead harmony in human being in all phases of living.

UHV language with STEM implementation will not only remove the fear of students in educational environment but also be a holistic and sustainable education approach and lead to holistic and sustainable development of technology, and production. It is possible to predict grades, behavior, and holistic and sustainable education by using Educational Data Mining techniques like classification and clustering by implementing UHV-STEM with EDM features. In the end, it is possible to alter the students' belief systems and modify their learning strategy from learning simply STEM to UHV-STEM. Since in the research both Science and Arts students, and faculties participated so both model UHVSTEM and UHVSTEAM are accurate for implementation. The use of new educational policies, their analysis, and the difficulties that students have previously encountered are all important for removing agitation from human life.

Predicting an educational model for a social, economic, sustainable, co-existential, holistic, and comprehensive approach to education is the main goal of this research. Educational Data Mining tools, techniques, and methods are having an impact on the educational environment, not only in terms of body-centric approaches like mark prediction, but also have an impact on self-centric approaches like behavior prediction, in line with the objectives of New Education Policies across the world, where course, marks and behavior prediction are the two main features of EDM. It is possible to forecast more accurately the inclusion of courses like Universal Human Values with STEM and NON-STEM by using Educational Data Mining and applying machine learning, artificial intelligence, and other techniques.

8. Conclusion and Future Work

The primary goal of this study is to assess the level of interest and perceived importance of Universal Human Values (UHV) among students in arts and science disciplines. The study aims to determine if there is a greater inclination among science students towards studying UHV as a mandatory subject or language. Currently, the global education model is predominantly focused on skill development, known as STEM education. While STEM education enhances practical skills, there is a lack of emphasis on understanding human relationships, leading to a sense of disconnect among students and families over time.

Through the analysis of collected data using K-Means clustering algorithms, this research underscores the importance of UHV in fostering understanding and meaningful relationships among individuals, ultimately contributing to happiness. The study advocates for the integration of UHV into STEM education to promote harmony in society. The proposed education model should evolve into UHVSTEM, emphasizing the importance of both Universal Human Values and STEM subjects

Certainly, AI-based STEM education and the integration of machine learning into every part of learning are excellent advancements for this era, but it also raises the question of whether these technical advances will result in comprehensive and sustainable development. Will this development not lose the temperament of students as well as all human being. Will normal prediction of students' performance be possible? Perhaps answer of all these questions will be a big NO to the world over.

Even it will improve disorder of the students as well as teachers also. Definitely for prediction, the methods, instruments, and strategies of Educational Data Mining through Machine Learning and AI will work out, and any type of disorder can be controlled by applying, comprehending, and learning the Universal Human Values with STEM (UHV-STEM) subjects in the form of courses, programs, and subjects in the educational system. Upon recognizing the significance of the student-centric, body-centric, and skill-centric teaching learning approaches, it becomes evident that the teacher-centric, self-centric, and value-centric approaches are equally important for all students. UHV-STEM implementation will also can control the uses of Artificial Intelligence (AI), that how much it has to be implemented at the level of Humanity. Future growth in the form of the investigation of UHV material with teachers and students alike may make it feasible to forecast student behavior based on the self-assessment batteries.

Funding: No funding was provided for the completion of this study.

Declarations

Conflict of Interest: The researcher declares that they have no conflict of interest.

References

- Abu Tair, M. M., & El-Halees, A. M. (2012). Mining educational data to improve students' performance: a case study. *International Journal of Information*, 2(2). http://hdl.handle.net/20.500.12358/25066 Corpus ID: 661468
- Adekitan, A. I., & Noma-Osaghae, E. (2019). Data mining approach to predicting the performance of first year student in a university using the admission requirements. *Education and Information Technologies, 24*, 1527-1543. https://doi.org/10.1007/s10639-018-9839-7
- Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD *international conference on Management of data* (pp. 207-216). https://doi.org/10.1145/170036.170072
- Araka, E., Oboko, R., Maina, E., &Gitonga, R. (2022). Using Educational Data Mining Techniques to Identify Profiles in Self-Regulated Learning: An Empirical Evaluation. *The International Review of Research in Open and Distributed Learning*, 23(1), 131–162. https://doi.org/10.19173/irrodl.v22i4.5401
- Arnold, J. C., Mühling, A., & Kremer, K. (2021). Exploring core ideas of procedural understanding in scientific inquiry using educational data mining. *Research in Science & Technological Education*, 1–21.https://doi.org/10.1080/02635143.2021.1909552
- 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. https://doi.org/10.1016/j.procs.2020.03.358
- Asif, R., Merceron, A., Ali, S. A., &Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113, 177–194.https://doi.org/10.1016/j.compedu.2017.05.007
- Babu, G., & Satya, S. (2022). Understanding the Inherent Interconnectedness and other Salient Characteristics of Nature crucial for Sustainability. *Environment, Development and Sustainability,* 1-13. https://doi.org/10.1007/s10668-022-02831-4
- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future

- visions. *Journal of educational data mining*, *1*(1), 3-17.
- Baker, R. S. J. D. (2010). Data mining for education. *International encyclopedia of education*, 7(3), 112-118.https://doi.org/10.1016/b978-0-08-044894-7.01318-x
- Buniyamin, N., bin Mat, U., & Arshad, P. M. (2015). Educational data mining for prediction and classification of engineering students achievement. *2015 IEEE 7th International Conference on Engineering Education (ICEED)*, 49–53.https://doi.org/10.1109/iceed.2015.7451491
- Cerda, P., Varoquaux, G., & Kégl, B. (2018). Similarity encoding for learning with dirty
- categorical variables. *Machine Learning*, *107*(8-10), 1477-1494.
- https://doi.org/10.1007/s10994-018-5724-2
- Devasia, T., Vinushree, T. P., &Hegde, V. (2016). Prediction of students performance using Educational Data Mining. *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, 91–95. https://doi.org/10.1109/sapience.2016.7684167
- Durairaj, M., &Vijitha, C. (2014). Educational data mining for prediction of student performance using clustering algorithms. *International Journal of Computer Science and Information Technologies*, 5(4), 5987–5991. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.567.8824&rep=rep1
- &type=pdf Corpus ID: 17278167

- Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2019).
- Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, 94, 335–343. https://doi.org/10.1016/j.jbusres.2018.02.012

- Gaur, R.R.,Sangal, R., & Bagaria, G.P.(2010). A Foundation course in Human Values and Professional Ethics. 2010, *Excel Books Private Limited*.
- Gupta, M. K., & Tiwari, M. (2022). Effect Of Universal Human Values On Male Prisoners At Mandoli Central Jail, Delhi, India. *Journal of Positive School Psychology*, 7236-7248.
- Ivančević, V., Knežević, M., Pušić, B., &Luković, I. (2014). Adaptive testing in programming courses based on educational data mining techniques. In *Educational data mining* (pp. 257–287). Springer. https://doi.org/10.1007/978-3-319-02738-8_10
- Jindal, R., & Borah, M. D. (2013). A survey on educational data mining and research trends. *International Journal of Database Management Systems*, 5(3), 53. https://doi.org/10.5121/ijdms.2013.5304
- Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British journal of applied science & technology*, *7*(4), 396-403. https://doi.org/10.9734/bjast/2015/14975
- Levin, N. A. (2021). Process Mining Combined with Expert Feature Engineering to Predict Efficient Use of Time on High-Stakes Assessments. *Journal of Educational Data Mining*, 13(2), 1-15. https://eric.ed.gov/?id=EJ1320549
- Liu, N., Zhang, B., Yan, J., Chen, Z., Liu, W., Bai, F., & Chien, L. (2005). Text representation: From vector to tensor. *Fifth IEEE International Conference on Data Mining* (ICDM'05), 4--pp.https://doi.org/10.1109/icdm.2005.144

- Merceron, A., &Yacef, K. (2005). Educational Data Mining: a Case Study. *AIED*, 467–474. https://www.semanticscholar.org/paper/Educational-Data-Mining%3A-a-Case-Study-Merceron-Yacef/9eb1b8cd6f7e379e505ac4f1ea5a95d16b8e259fCorpus ID: 10949661
- Mumcu, F., Uslu, N. A., & Yıldız, B. (2023). Teacher development in integrated STEM education: Design of lesson plans through the lens of computational thinking. *Education and Information Technologies*, *28*(3), 3443-3474. https://doi.org/10.1007/s10639-022-11342-8
- Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for Educational Data Mining models: A case study in affect detection. *British Journal of Educational Technology*, 45(3), 487–501.https://doi.org/10.1111/bjet.12156
- Peña-Ayala, A., & Cárdenas, L. (2014). How educational data mining empowers state policies to reform education: The Mexican case study. In *Educational Data Mining* (pp. 65–101). Springer. https://doi.org/10.1007/978-3-319-02738-8_3
- Prada, M. A., Dom\'\inguez, M., Vicario, J. L., Alves, P. A. V., Barbu, M., Podpora, M., Spagnolini, U., Pereira, M. J. V., &Vilanova, R. (2020). Educational data mining for tutoring support in higher education: a web-based tool case study in engineering degrees. *IEEE Access*, 8, 212818–212836. https://doi.org/10.1109/access.2020.3040858
- Raval, U. R., & Jani, C. (2016). Implementing & improvisation of K-means clustering algorithm. *International Journal of Computer Science and Mobile Computing*, *5*(5), 191-203.
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146.https://doi.org/10.1016/j.eswa.2006.04.005
- Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics* Part C: Applications and Reviews, 40(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532
- Saa, A. A. (2016). Educational data mining & students' performance prediction. *International Journal of Advanced Computer Science and Applications*, 7(5).https://doi.org/10.14569/ijacsa.2016.070531
- Scheuer, O., & McLaren, B. M. (2012). Educational data mining. Encyclopedia of the Sciences of Learning, 1075–1079.https://doi.org/10.1007/978-1-4419-1428-6_618
- Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., &Gasevic, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics*, 42(1), 85–106. https://doi.org/10.3102/1076998616666808
- Wright, G. W., & Delgado, C. (2023). Generating a framework for gender and sexual diversity-inclusive STEM education. *Science Education*, *107*(3), 713-740. https://doi.org/10.1002/sce.21786
- Xia, X. (2020). Clustering Analysis of Interactive Learning Activities Based on Improved
- BIRCH Algorithm. *arXiv preprint arXiv:2010.03821*.
- Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: a systematic review
- from 2011 to 2021. *International Journal of STEM Education*, 9(1), 1-20. https://doi.org/10.1186/s40594-022-00377-5

- Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11.
- Zaffar, M., Hashmani, M. A., & Savita, K. S. (2017). Performance analysis of feature selection algorithm for educational data mining. *2017 IEEE Conference on Big Data and Analytics*

(ICBDA), 7–12. https://doi.org/10.1109/icbdaa.2017.8284099

BIOGRAPHIES OF AUTHORS

	**Dheeraj Kumar Singh** is a Ph.D. Research scholar in the Department of Computer Science, School of Technology, Doon University, Dehradun. In addition to serving as an AICTE volunteer for UHV work, he co-facilitates, observes, and helps out at numerous UHV events in India. His MCA degree is from UPTU, Lucknow (U.P.), and his M.Tech. in information technology is from GGSIP University, Delhi. In addition to being a member of numerous academic and administrative committees, he has been in teaching from 2004. His expertise includes working with educational data mining, programming languages, machine learning, digital marketing, patents, trademarks, and copyrights. He can be contacted at email: dheerajsingh.phd@gmail.com
	**Narender Kumar** is Ph.D. from Indian Institute of Technology (I.I.T.), Roorkee. He is currently working as an Associate Professor in the Department of Computer Science, School of Technology, Doon University, Dehradun, since 2021. Before joining to Doon University, He has worked as Assistant Professor in the Department of Computer Science and Engineering, H.N.B Garhwal University, Srinagar Garhwal from 2012 to 2021. His main research work focuses on Optimization Algorithms, Blockchain Technology, Data Mining, Machine Learning, Genetic Algorithm, PSO and Digital Image Processing. He has 14 years of teaching experience and 10 years of Research Experience.He can be contacted at email: narenrawal@gmail.com