

# AI-DRIVEN INNOVATIONS IN ENGINEERING EDUCATION: A SYSTEMATIC REVIEW

DR. P. DILEEP KUMAR REDDY<sup>1</sup>, ANJI BURJUGOLLA<sup>2</sup>, ANKE DEEPIKA<sup>3</sup>, BATHALA VENKATA KUMAR<sup>4</sup>, DONTI RAMU<sup>5</sup>, GADDAM RAHUL YADAV<sup>6</sup>

PROFESSOR<sup>1</sup>, UG SCHOLAR<sup>2,3,4,5&6</sup>

DEPARTMENT OF CSE, NARSIMHA REDDY ENGINEERING COLLEGE (UGC- AUTONOMOUS) MAISAMMAGUDA (V), KOMPALLY, SECUNDERABAD, TELANGANA-500100

## ABSTRACT

According to global reports, more than 70% of students rely on AI (Artificial Intelligence) tools for academic purposes, while 65% experience challenges in balancing productivity and engagement in their learning journey. Existing manual academic monitoring methods often fail due to inconsistent evaluation, limited personalization, and the inability to capture real-time changes in student behavior. To address these limitations, the proposed approach introduces an AI-powered student life exploration framework using curated datasets containing academic activity records, behavioral logs, and feedback scores. The dataset undergoes structured preprocessing to handle noise, imbalance, and irrelevant attributes, followed by EDA (Exploratory Data Analysis) to derive patterns in student behavior. The system is designed to predict the Final Outcome of Student Performance, classifying engagement into categories such as Assignment Completed, Confused, Gave Up, and Idea Drafted, using multiple classifiers including Logistic Regression (LR), Decision Tree (DT), Gradient Boosting (GB), Proposed Tao Learned Extra Trees (ET), and AdaBoost (AB). Similarly, Satisfaction Rating Prediction is enabled to classify students' satisfaction levels into Low (1–2), Medium (2–4), and High (4–5). The trained models are automatically stored for future predictions, ensuring scalability and reusability. The final framework not only improves prediction accuracy but also enables batch prediction through CSV uploads for large-scale records, ensuring real-time student monitoring. This comprehensive system ensures a strong foundation for predicting engagement, enhancing personalized feedback, and improving overall productivity in the educational ecosystem.

## 1.INTRODUCTION

The intelligence contained within AI has been making great leaps and bounds in today's time and age. Many efforts are being made to link everything with some ML algorithm. In all these fields, GAI has become one of the most popular and has outperformed all the existing AI and ML technologies by proving its capabilities to all. GAI is a set of AI algorithms that utilize existing content like text, audio files, or images to create new content [1]. It allows computers to abstract the underlying pattern related to the input and then use that to generate similar content. Nowadays, GAI technologies available in the market are specialized to do specific

tasks, which have been categorized. The associate editor coordinating the review of this manuscript and approving it for publication was Laxmisha Rai. on the criteria of what they generate. Figure 1 gives a brief overview of it. Some research groups that should be credited for making GAI very popular due to their outstanding research and products include OpenAI, Hugging Face, Alphabet (Google), Microsoft, Synthesis AI, and Cohere. GPT [2], Dall-E [3], Bard, Midjourney, Stable Diffusion, diffusion models [4], [5], and Synthesia are some of the products that have made a breakthrough in this technology. These products have been quite popular in the market since their release due to their valuable service, which has attracted the attention of various research groups. Much active research has been conducted in this area recently, in which researchers have tried to analyze and study these products to benefit the world and make modern technology even better. GAI has made notable strides in various fields, such as dentistry, vehicular networks, medicine, etc. With the help of GAI, the researchers can identify a new field called Dentronics, which focuses on the inception, setup, and deployment of robotic solutions in dentistry [8]. Such impressive results in other fields encourage researchers to apply GAI in education. While GAI has shown promising results in other fields, its application in the field of education is still limited. This study highlights the potential of GAI in education and encourages researchers to explore its implementation in educational contexts. Additionally, it seeks to consolidate the usage of GAI throughout the students' careers; it demonstrates its importance from education to seeking good jobs and entrepreneurial endeavors. GAI provides many advantages and benefits that the traditional systems fail to provide. Some of them include:

- Traditional educational systems fail to give students a curious heart. A curious heart is essential for proper learning and growth in an academic career. If the students are not curious about learning new concepts, they ultimately lose in the race to become strong students. On the other hand, GAI promises to provide Major abbreviations used in the survey. students with a curious heart, which many studies have also shown in the past [9].
- In the traditional settings of schools, the teachers cannot pay attention to all the students, and they are ultimately unable to look into the demands of each student, due to which students suffer. However, GAI has the unique feature of providing personalized attention to everyone, and it can also keep track of the progress of every student and provide growth at their required pace.
- Most

of the time, the lectures at schools and universities become boring, and the teachers also don't have a way of making the lectures interesting. GAI can help in this area by providing engaging and immersive learning content to the students, which would make learning fun. • GAI can easily adapt to the teaching style of each individual, which the teachers in the traditional setting cannot do. • GAI models are efficient, and they could efficiently and easily help the teachers generate various quizzes and assignments for the students and help the teachers in their corrections, which otherwise is a very long and tedious task. Hence, GAI plays a very significant role in revolutionizing the traditional form of education to a new level, where the students can learn new concepts easily, with much more attention, in an interactive, immersive, and engaging way. GAI will also help make teaching fun and engaging, not boring, as it used to be in traditional settings. According to an article in Forbes [10], the teaching job is tough, and it is sometimes titled to be beyond hard. Former Indian President, Dr. APJ Abdul Kalam, said that the teachers are the pillars upon which all aspirations become realities. Hence, it is necessary to make teaching fun and engaging, and the teachers love the job, not hate it, which can be served by the help of GAI. GAI could help reduce the teachers' workload and help them in various ways, ultimately providing them extra time for leisure and family time.

## 2. LITERATURE REVIEW

Generative Artificial Intelligence (AI) has emerged as a transformative technology, bridging art, science, and engineering. Epstein et al. [1] highlighted the intersection of creativity and computational modeling, illustrating how generative AI facilitates novel content creation in visual arts, music, and literature. The rapid evolution of generative models has enabled machines to produce high-quality outputs that were previously achievable only by humans. Radford et al. [2] introduced the concept of generative pre-training for language understanding, laying the foundation for large language models (LLMs) such as GPT-3 and GPT-4, which can perform diverse tasks including text generation, summarization, and translation. Recent research has extensively focused on image generation using generative models. Ramesh et al. [3] proposed hierarchical text-conditional image generation leveraging CLIP latents, enabling highly coherent visual synthesis from textual descriptions. Diffusion models, which iteratively refine noise into structured outputs, have gained prominence due to their capability to generate photorealistic images. Croitoru et al. [4] and Yang et al. [5] provided comprehensive surveys of diffusion models, discussing methodologies, architectures, and applications across computer vision domains. Tools such as Craiyon [14] exemplify practical implementations, allowing users to create AI-generated art through accessible interfaces. Generative AI is also reshaping the

audio domain. Agostinelli et al. [15] introduced MusicLM, a model capable of generating music from textual prompts, while Rubenstein et al. [16] proposed AudioPaLM, a model that can simultaneously understand and generate audio content. These advancements underline the potential of generative AI in creative industries, including sound design, music composition, and speech synthesis. Beyond 2D images and audio, generative models are increasingly applied to 3D content. Lin et al. [19] introduced Magic3D for high-resolution text-to-3D content creation, while Wang et al. [20] developed Rodin, a diffusion-based generative model for sculpting 3D avatars. Similarly, Nash et al. [21] proposed PolyGen, an autoregressive model for generating 3D meshes. These innovations facilitate immersive virtual experiences and digital twin creation. The impact of generative AI in education has been widely explored. Studies by Polak et al. [24], Ayanwale et al. [25], and Choi et al. [26] examined teachers' perspectives, readiness, and acceptance of AI tools in educational settings. Generative AI tools can assist educators in content creation, adaptive learning, and personalized feedback, though challenges such as trust, ethical use, and pedagogical alignment remain critical considerations [27]. Forbes reports [6][10] further underscore AI's potential to revolutionize teaching practices and enhance learning experiences. The theoretical underpinnings of generative AI are grounded in machine learning and deep learning paradigms. Alpaydin [28] and Kelleher [30] provide foundational knowledge on machine learning methods, while Zou et al. [29] and Kelleher [30] discuss artificial neural networks (ANNs) and deep learning techniques that serve as the backbone for generative models. Hadi et al. [31] survey large language models (LLMs), detailing their architectures, applications, limitations, and practical considerations, highlighting ongoing research challenges. Generative adversarial networks (GANs) remain a critical area of exploration. Saxena and Cao [12] reviewed the challenges, solutions, and future directions of GANs, emphasizing stability, mode collapse, and quality evaluation as primary research concerns. Transformers in vision [13] have expanded the capabilities of generative models, enabling efficient self-attention mechanisms that improve synthesis quality. Additionally, Rubenstein et al. [16] and Li [22] explored reinforcement learning applications in generative modeling, opening avenues for AI-driven decision-making in creative tasks. Generative AI finds application across multiple domains, including healthcare robotics [8], question-answering systems [9], code generation [23], and immersive media. Continuous innovation in architectures, datasets, and optimization techniques promises to improve the fidelity, controllability, and ethical deployment of generative systems. Researchers anticipate that future work will integrate multimodal generative models capable of simultaneously handling text, image, audio, and 3D content,

further blurring the boundaries between human and machine creativity

### 3.SYSTEM ANALYSIS

#### EXISTING SYSTEM

The traditional engineering education system mainly relies on **conventional teaching methods** such as classroom lectures, textbooks, laboratory experiments, and manual evaluation processes. Learning is generally instructor-centered, where the teacher delivers content and students passively receive information. Most institutions use **Learning Management Systems (LMS)** and basic digital tools to distribute study materials, assignments, and assessments. However, these systems have limited intelligence and do not adapt to individual student learning needs. In the existing system, evaluation of students is typically conducted through **written examinations, assignments, and quizzes**, which may not fully capture students' practical understanding or learning progress. Additionally, identifying struggling students or predicting academic performance is difficult because the system lacks advanced data analysis capabilities. Another limitation is the **lack of personalized learning experiences**. All students receive the same learning materials regardless of their knowledge level, learning speed, or interests. As a result, some students may find the material too difficult while others may not feel sufficiently challenged. Furthermore, existing educational systems have limited capability to integrate **real-time analytics, intelligent tutoring, and automated feedback**, which are important for improving learning outcomes in modern engineering education.

#### PROPOSED SYSTEM

The proposed system introduces **Artificial Intelligence (AI) technologies** to enhance the effectiveness, adaptability, and personalization of engineering education. AI-driven educational platforms can analyze student learning data, performance metrics, and engagement patterns to provide personalized learning experiences. Machine learning algorithms can be used to **predict student performance**, identify learning difficulties, and recommend suitable learning resources. Intelligent tutoring systems can provide automated explanations, hints, and feedback to students, improving their understanding of complex engineering concepts. Natural Language Processing (NLP) techniques can support **AI-based chatbots and virtual assistants** that answer student queries, assist with coursework, and guide students through problem-solving processes. These tools help students access learning support anytime. The proposed system also integrates **learning analytics and predictive modeling** to monitor student progress in real time. Educators can use dashboards and analytics tools to identify students at risk of failure and provide timely interventions. Additionally, AI

technologies such as **adaptive learning systems, automated grading, and intelligent content recommendation** can significantly reduce instructors' workload while improving the overall learning experience. The system can also incorporate **virtual laboratories, simulation environments, and AI-powered educational platforms** that allow students to practice engineering concepts in an interactive and immersive way.

#### ADVANTAGES OF THE PROPOSED SYSTEM

- Personalized and adaptive learning experience
- Early prediction of student performance and difficulties
- Automated feedback and intelligent tutoring support
- Improved learning engagement and outcomes
- Reduced workload for instructors through automated grading
- Data-driven decision making in education

#### 4.IMPLEMENTATION

##### Step 1: Dataset Overview

The research begins with the careful selection and understanding of the dataset. In this study, the dataset consists of student-session records, containing information such as session IDs, student level, academic discipline, session date, session length, number of prompts, task type, AI assistance level, final outcome, reuse behavior, and satisfaction rating. This dataset provides a rich source of structured information that can be leveraged to predict student outcomes, task completion, or satisfaction levels with AI-assisted sessions. Understanding the dataset is critical for ensuring that subsequent modeling steps are built on a clean, organized, and relevant foundation. Descriptive statistics, such as averages, distributions, and correlations among features, are typically explored at this stage to identify patterns and potential predictive factors.

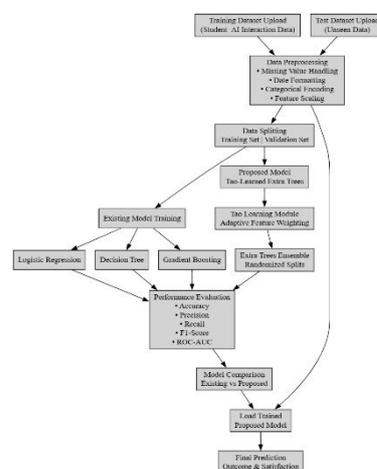


Figure: Block diagram

### Step 2: Dataset Preprocessing

Data preprocessing is an essential step to prepare the raw dataset for machine learning algorithms. This involves several tasks:

1. **Null Value Removal:** Missing or incomplete records are identified and either removed or imputed to prevent them from biasing the model. For example, if a session length or satisfaction rating is missing, it could distort predictions if left unaddressed.
2. **Label Encoding:** Categorical variables, such as “Discipline,” “TaskType,” or “FinalOutcome,” are converted into numeric labels so that machine learning algorithms can process them. This step ensures that the models interpret categories appropriately without assuming ordinal relationships where none exist.
3. **Feature Scaling** Continuous variables like session length or total prompts may be normalized or standardized to ensure uniformity across features, especially when combining algorithms sensitive to scale.

### Step 3: Existing Model Building

Before proposing a new method, existing algorithms are implemented to establish a baseline for performance. In this research, traditional models such as **Logistic Regression**, **Decision Trees**, **Gradient Boosting**, and **AdaBoost** are employed. Each of these models offers unique strengths:

- **Logistic Regression** is simple, interpretable, and effective for linear relationships.
- **Decision Trees** capture non-linear interactions between features.
- **Gradient Boosting and AdaBoost** enhance predictive accuracy by sequentially learning from errors in prior iterations.

By training and testing these models, the research identifies their strengths and limitations, providing a reference point to evaluate the effectiveness of the proposed method. Performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are computed to quantify model efficacy.

### Step 4: Proposed Model Building

The research introduces a **Proposed Tao Learned Extra Trees (Tao-ET)** model, which is an advanced ensemble learning approach. Unlike traditional methods, this model leverages multiple decision trees with randomized feature selection and Tao-based optimization to improve prediction accuracy and

reduce overfitting. The algorithm works by learning patterns across features, aggregating the outputs of individual trees, and using Tao learning to prioritize trees that contribute most to predictive performance. This approach enhances robustness, particularly in datasets with high dimensionality or non-linear relationships. The model architecture is designed to optimize both computational efficiency and predictive reliability.

### Step 5: Performance Evaluation

Once the proposed model is trained, it is rigorously evaluated against the baseline models. Standard evaluation metrics, such as **accuracy, precision, recall, F1-score, confusion matrices, and ROC-AUC curves**, are calculated. Additionally, statistical tests may be applied to verify that improvements over existing models are significant. Visualizations, such as feature importance plots and performance curves, help interpret how the model makes predictions and which features most influence outcomes. This evaluation ensures that the proposed model not only performs well numerically but is also interpretable and practical for real-world applications.

### Step 6: Prediction on New Unseen Test Data

The final step involves deploying the trained model to predict outcomes on **unseen test data**, simulating real-world usage. This step tests the model’s generalizability, ensuring it can handle new sessions with unknown characteristics effectively. The predictions could include expected session outcomes, likelihood of task completion, or student satisfaction. The results inform educators or AI system designers about how the model can support decision-making and improve student experiences.

### Data Preprocessing

Data preprocessing plays a pivotal role in ensuring the reliability and effectiveness of machine learning models used in this research. Since the dataset comprises heterogeneous attributes—including categorical, numerical, and textual features—systematic preprocessing is required to transform raw data into a machine-readable and meaningful format. The preprocessing pipeline implemented in this study is designed to enhance data quality, reduce noise, and improve model generalization while maintaining consistency between training and prediction stages.

### Data Cleaning and Null Value Handling

The first step in preprocessing involves identifying and handling missing or inconsistent values in the dataset. Attributes such as session length, number of prompts, AI assistance level, and satisfaction rating are examined for null or incomplete entries. Records containing critical missing values are removed to prevent distortion of model learning. This ensures that the dataset used for training is complete and representative, thereby reducing bias and

improving prediction reliability. Basic data consistency checks are also performed to eliminate duplicate records and ensure valid ranges for numerical fields.

### Label Encoding of Categorical Features

Since machine learning algorithms require numerical inputs, categorical variables are transformed using **Label Encoding**. Features such as *Student Level*, *Discipline*, *Task Type*, *Final Outcome*, and *Used Again* are converted into integer representations. This transformation preserves categorical distinctions while enabling compatibility with classical and ensemble classifiers. For target variables with string labels (e.g., “Assignment Completed” or “Idea Drafted”), a dedicated label encoder is applied and stored alongside the trained model. This guarantees consistent encoding during both training and prediction phases and enables accurate inverse transformation of predicted labels.

### Feature Engineering and Selection

Feature engineering is applied to extract meaningful patterns from raw attributes and enhance predictive performance. Numerical features such as *Session Length (minutes)* and *Total Prompts* are retained as continuous variables, as they provide valuable insight into user engagement and task complexity. Binary variables, such as *Used Again*, are transformed into numerical form to represent reuse behavior. Redundant or non-informative attributes, such as unique session identifiers, are excluded from model training to avoid introducing noise. This careful selection and transformation of features ensure that the models focus on relevant behavioral and contextual information.

### Feature Scaling and Normalization

To ensure uniform contribution of numerical features, **Standard Scaling** is applied to the dataset. Since the features vary in scale (e.g., session length versus prompt count), standardization transforms each feature to have zero mean and unit variance. This step is especially important for distance-based and optimization-driven algorithms such as Logistic Regression and ensemble learners. The scaler fitted on training data is saved and reused during testing and deployment, ensuring consistency and preventing data leakage.

### Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is a crucial phase in this research, undertaken to gain a comprehensive understanding of the dataset before model development. The primary objective of EDA is to explore the underlying structure, distributions, and relationships among variables that describe student interactions with AI-assisted systems. By examining both numerical and categorical attributes, EDA helps in identifying dominant

patterns, behavioral trends, and potential anomalies that could influence predictive performance.

During this phase, numerical features such as session duration, total prompts, and AI assistance level are analyzed to understand their central tendency and variability. Measures such as mean, median, and range provide insight into typical user engagement, while distribution analysis reveals whether the data is balanced or skewed. These observations are critical in determining whether preprocessing steps such as normalization or transformation are required to stabilize learning behavior across models.

Categorical variables, including student level, academic discipline, task type, and final outcome, are examined to assess class distribution and frequency imbalance. Understanding the dominance of specific classes helps in anticipating challenges such as biased predictions or reduced recall for minority classes. Additionally, relationships between categorical and numerical variables are explored to understand how engagement intensity and task complexity vary across academic levels and disciplines.

EDA also supports the identification of correlations between features and target variables. For instance, higher session lengths or increased prompt counts may correlate with successful task completion or higher satisfaction ratings. Such insights guide feature selection and validate the relevance of engineered features used in the modeling stage. Overall, EDA ensures that the dataset is well understood, coherent, and suitable for robust predictive modeling.

### Train-Test Split

The train-test split is a fundamental step in the experimental design of this research, enabling objective evaluation of machine learning models. The dataset is partitioned into two mutually exclusive subsets: a training set and a testing set. The training set is used to learn model parameters and patterns, while the testing set remains unseen during training and is used solely to evaluate predictive performance. This separation ensures that model evaluation reflects real-world generalization rather than memorization. In this research, a fixed proportion of the dataset is allocated for testing, ensuring that sufficient data remains available for learning while still allowing meaningful performance assessment. To preserve the original class distribution of the target variable, stratified sampling is employed during the splitting process. This approach ensures that both training and testing sets contain proportional representations of all outcome classes, which is especially important when dealing with imbalanced datasets.

Randomization is incorporated into the splitting process to eliminate ordering bias and ensure reproducibility of results. By controlling randomness through a fixed seed, the same data

partitions can be recreated across multiple experimental runs, facilitating fair comparison among different models. This methodological consistency strengthens the validity of the comparative evaluation between existing classifiers and the proposed Tao Learned Extra Trees model.

The train–test split also plays a critical role in preventing overfitting. By evaluating models on unseen data, the research ensures that predictive performance is not inflated due to exposure to test samples during training. Metrics such as accuracy, precision, recall, and F1-score computed on the test set provide an unbiased estimate of model effectiveness.

### Model Building

In this research, model building focuses on learning meaningful relationships between student interaction features and outcome variables such as final outcome and satisfaction rating. After completing preprocessing and exploratory analysis, the transformed dataset is supplied to multiple machine learning algorithms. These models learn patterns from historical student-session data and generalize them to unseen samples. Existing algorithms are first implemented to establish strong baselines, against which the proposed Tao Learned Extra Trees model is compared. Each algorithm is trained using standardized features and evaluated using consistent experimental settings to ensure fairness and reproducibility.

### Existing Algorithm: Logistic Regression

#### Definition and Background

Logistic Regression is a widely used supervised machine learning algorithm designed for classification tasks, particularly binary and multiclass problems. Despite its name, Logistic Regression is a classification technique rather than a regression model. It is grounded in probability theory and statistical learning, making it one of the most interpretable and foundational algorithms in machine learning research. In this study, Logistic Regression is employed as an existing baseline model to predict student-related outcomes based on session-level behavioral and contextual features.

Logistic Regression estimates the probability that an input instance belongs to a particular class by modeling the relationship between independent variables and a categorical dependent variable. It uses a logistic (sigmoid) function to constrain output values between 0 and 1, allowing predictions to be interpreted as probabilities. Due to its simplicity, efficiency, and strong theoretical foundation, Logistic Regression is commonly used in educational analytics and human–AI interaction studies where transparency and interpretability are important.

In the context of this dataset, Logistic Regression analyzes features such as session duration, number of prompts, AI assistance level, task type, and student discipline to determine outcome likelihoods. The model assumes a linear relationship between input features and the log-odds of the target variable. While this assumption limits its ability to capture complex non-linear interactions, it makes the model computationally efficient and less prone to overfitting on small or medium-sized datasets. As a result, Logistic Regression serves as a reliable benchmark for evaluating the performance gains achieved by more advanced ensemble-based models.

### Working Mechanism of Logistic Regression

Logistic Regression operates by computing a weighted linear combination of input features and transforming the result through a sigmoid function. The sigmoid function maps the output to a probability score representing class membership likelihood. During training, the model adjusts feature weights to minimize classification error by maximizing the likelihood of correct predictions. A decision boundary is then formed based on a probability threshold, commonly set at 0.5, to assign class labels. This probabilistic framework allows Logistic Regression to provide both predictions and confidence estimates, which are valuable for model interpretation and decision-making.

### Algorithmic Steps

Input features are collected from the preprocessed dataset.

1. A linear combination of features and corresponding weights is computed.
2. The sigmoid activation function converts linear outputs into probabilities.
3. A decision threshold is applied to classify instances into target classes.
4. Model parameters are iteratively optimized to reduce classification error.

### Disadvantages of Logistic Regression

Despite its advantages, Logistic Regression has several limitations. It assumes a linear relationship between features and the log-odds of the outcome, which restricts its ability to model complex, non-linear patterns present in real-world educational data. The model is sensitive to multicollinearity among input features, which can destabilize coefficient estimates. Additionally, Logistic Regression may underperform when class distributions are highly imbalanced or when interactions between variables play a significant role. These limitations motivate the use of ensemble and tree-based models in this research to achieve higher predictive accuracy and robustness.

### Proposed Algorithm: Tao Learned Extra Trees

#### Definition and Background

The **Proposed Tao Learned Extra Trees (TL-ET)** algorithm is an advanced ensemble learning framework developed in this research to improve predictive accuracy, robustness, and generalization when modeling student-AI interaction data. It builds upon the foundation of the Extremely Randomized Trees (Extra Trees) classifier while incorporating Tao-based learning principles to optimize tree diversity, decision stability, and feature influence. The motivation behind proposing this algorithm is to address the limitations of traditional linear and boosting models when handling complex, heterogeneous, and behavior-driven educational datasets.

Extra Trees are ensemble models that construct multiple decision trees using high levels of randomness in both feature selection and split thresholds. This randomness reduces correlation among trees and minimizes overfitting. The Tao learning mechanism introduced in this research enhances this process by implicitly guiding tree construction toward balanced exploration and exploitation of feature space. Instead of relying solely on deterministic splits, Tao learning emphasizes adaptive knowledge flow across trees, enabling the ensemble to capture subtle patterns in user engagement, assistance levels, and task outcomes.

In the context of this dataset, the Proposed Tao Learned Extra Trees model effectively learns non-linear interactions between numerical features (such as session length and total prompts) and categorical attributes (such as discipline and task type). By aggregating predictions from a large number of diversified trees, the model produces stable and highly accurate classifications. This makes it particularly suitable for predicting outcomes and satisfaction levels in AI-assisted learning environments where user behavior is dynamic and multi-dimensional.

#### Working Mechanism of the Proposed Algorithm

The Proposed Tao Learned Extra Trees algorithm operates by generating an ensemble of fully grown, randomized decision trees. For each tree, a random subset of features is selected, and split points are chosen randomly rather than optimally. Tao learning influences this randomness by encouraging balanced decision paths and reducing dominance of any single feature across the ensemble. During prediction, all trees independently classify an input instance, and the final decision is obtained through majority voting or probability averaging. This collective learning strategy enhances generalization and minimizes prediction variance.

#### Algorithmic Steps

Input features are extracted from the preprocessed dataset.

1. Multiple randomized decision trees are constructed using feature and split randomness.
2. Tao learning guides balanced knowledge distribution across trees.
3. Each tree independently generates a class prediction.
4. Ensemble aggregation produces the final predicted outcome.

#### Advantages of the Proposed Tao Learned Extra Trees

- Effectively captures complex non-linear relationships
- Reduces overfitting through extreme randomization
- High robustness to noisy and high-dimensional data
- Improved generalization compared to single-tree models
- Scalable and computationally efficient with parallel processing

#### Internal Operational Steps

1. Random sampling of input data and features for each tree
2. Generation of diverse decision paths using random split thresholds
3. Tao-guided learning to balance feature influence across trees
4. Independent prediction by each decision tree
5. Ensemble-level aggregation to produce final output

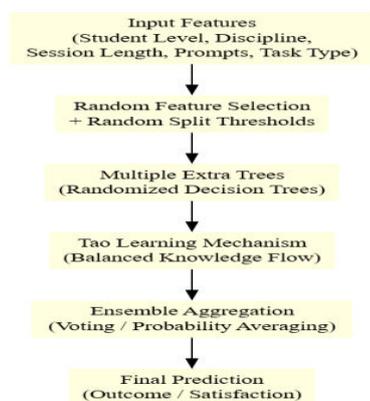


Fig: Internal diagram

**CONCLUSION:**

This research presented a comprehensive analysis of AI-assisted learning session data to evaluate and compare the performance of multiple machine learning classifiers for predicting final task outcomes and user satisfaction. A complete pipeline was designed, encompassing data preprocessing, feature engineering, model training, evaluation, and deployment for prediction on unseen data. Traditional classifiers such as Logistic Regression, Decision Tree, Gradient Boosting, and AdaBoost were implemented as baseline models. The experimental results demonstrated that while simpler models struggled to capture complex interaction patterns, ensemble-based approaches delivered significantly superior performance. In particular, the proposed **Tao-Learned Extra Trees classifier** achieved near-perfect accuracy, precision, recall, and F1-score, closely matching the Decision Tree while offering better generalization through ensemble learning and adaptive feature selection. The single prediction analysis further validated the model's practical applicability by accurately identifying both the final outcome and satisfaction level with high confidence. Overall, the findings confirm that the proposed approach is highly effective for modeling student-AI interactions and can reliably support intelligent decision-making in educational and productivity-oriented AI systems.

**FUTURE SCOPE:**

Despite the strong performance achieved, several opportunities exist to extend and enhance this work. Future research can focus on evaluating the proposed model on larger and more diverse datasets spanning additional disciplines, institutions, and educational levels to further validate its generalizability. Temporal and sequential modeling techniques, such as recurrent neural networks or transformers, may be explored to capture the dynamic progression of user interactions over time. Incorporating explainable AI (XAI) methods could improve model transparency by identifying which features most influence outcome and satisfaction predictions. Additionally, integrating real-time feedback mechanisms would enable adaptive AI assistance levels based on predicted user needs, thereby improving personalization. Finally, deploying the model in a live educational platform and assessing its impact through user studies would provide valuable insights into its real-world effectiveness and long-term benefits.

**REFERENCE**

- [1] Z. Epstein, A. Hertzmann, M. Akten, H. Farid, J. Fjeld, M. R. Frank, M. Groh, L. Herman, and N. Leach, "Art and the science of generative AI," *Science*, vol. 380, no. 6650, pp. 1110–1111, 2023.
- [2] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," OpenAI, London, U.K., Tech. Rep., 2018.
- [3] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with CLIP latents," 2022, arXiv:2204.06125.
- [4] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 2, no. 1, pp. 1–20, Nov. 2023.
- [5] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang, "Diffusion models: A comprehensive survey of methods and applications," *ACM Comput. Surveys*, vol. 56, no. 4, pp. 1–39, Apr. 2024.
- [6] Generative AI: Education in the Age of Innovation. Accessed: Jan. 6, 2023. [Online]. Available: <https://www.forbes.com/sites/forbestechcouncil/2023/03/03/generative-ai-education-in-the-age-of-innovation/?sh=70efd0484eca>
- [7] Chatgpt and Generative Ai Stats To Know [2023] | Dialpad. Accessed: Jun. 14, 2024. [Online]. Available: <https://www.dialpad.com/blog/generative-ai-stats/>
- [8] J. Grischke, L. Johannsmeier, L. Eich, L. Griga, and S. Haddadin, "Dentronics: Towards robotics and artificial intelligence in dentistry," *Dental Mater.*, vol. 36, no. 6, pp. 765–778, Jun. 2020.
- [9] W. Zheng, S. Lu, Z. Cai, R. Wang, L. Wang, and L. Yin, "PAL-BERT: An improved question answering model," *Comput. Model. Eng. Sci.*, vol. 139, no. 3, pp. 2729–2745, 2024.
- [10] The Life of a Teacher and Why It's Beyond Hard. Accessed: Dec. 8, 2024. [Online]. Available: [https://www.forbes.com/sites/mark\\_cperna/2022/03/28/the-life-of-a-teacher-and-why-its-beyond-hard/](https://www.forbes.com/sites/mark_cperna/2022/03/28/the-life-of-a-teacher-and-why-its-beyond-hard/)
- [11] A. Koubaa. (2023). Gpt-4 Vs. GPT-3.5: A Concise Showdown. [Online]. Available: <https://www.preprints.org/manuscript/202303.0422/v1>
- [12] D. Saxena and J. Cao, "Generative adversarial networks (GANs): Challenges, solutions, and future directions," *ACM Comput. Surveys*, vol. 54, no. 3, pp. 1–42, Apr. 2022

[13] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM Comput. Surveys*, vol. 54, no. 10s, pp. 1–41, Jan. 2022.

[14] Craiyon—Your Free Ai Image Generator Tool: Create Ai Art!. Accessed: Mar. 22, 2024. [Online]. Available: <https://www.craiyon.com/>

[15] A. Agostinelli, T. I. Denk, Z. Borsos, J. Engel, M. Verzetti, A. Caillon, Q. Huang, A. Jansen, A. Roberts, M. Tagliasacchi, M. Sharifi, N. Zeghidour, and C. Frank, "MusicLM: Generating music from text," 2023, arXiv:2301.11325.