Learning style based learner classifier using HistGradientBoosting

By

Maheswari R J
Department of Computer Science and Engineering
220701155

Guide

Dr. V.Auxilia Osvin Nancy., M.Tech, Ph.D.,
SUPERVISOR, Assistant Professor
Department of Computer Science and Engineering

Introduction

PROBLEM STATEMENT SOLUTION AND THE IMPACT

- E-learning often ignores individual learning preferences.
- One-size-fits-all content reduces learning effectiveness.
- Learner engagement drops without personalized delivery.

- Classification of learners into VARK styles
- Use a 15-question quiz to gather learner behavior data.
- Apply HistGradientBoostingClassifier to classify learning styles.

- Categories based on VARK: Visual, Auditory, Reading/Writing, Kinesthetic.
- Deploy model via a Flask REST API for real-time use.
- Enable personalized content delivery in digital platforms.

Literature Survey

Understanding Learning Styles & Traditional Limitations

Fleming's VAK Model (2006):
Categorizes learners into Visual, Auditory, or Kinesthetic based on how they best absorb and retain information.

Limitations of Traditional Self-Assessments:
Often rely on subjective responses; results may be inconsistent and not reflect real-world behavior or evolving learning needs.

Role of Machine Learning (ML):
ML offers the ability to analyze large datasets and identify learning preferences through user interaction patterns, enabling dynamic, adaptive learning environments.

Rising Need for Personalization:
With the shift toward learner-centric platforms, static assessments fall short in supporting real-time personalized learning.

Literature Survey

Machine Learning Approaches to Learning Style Prediction

Decision Trees (Ali & Al-Khanjari, 2020):
Used for predicting learning styles based on quiz responses; easy to interpret, making them practical for educational applications.

Support Vector Machines – SVM (Xu et al., 2017):
Superior accuracy in classifying subtle differences in learner behavior; effective in high-dimensional spaces with complex patterns.

Ensemble Learning Methods (Zhang et al., 2020):
Techniques like HistGradientBoostingClassifier combine multiple models to improve prediction accuracy, especially with noisy or incomplete data.

Real-Time Adaptation (Gibbons et al., 2018):
ML-powered systems adapt to user feedback and performance over time, modifying content and delivery methods to suit changing learner needs.

Literature Survey

Advancements, Challenges & Future Scope

Multimodal Data Integration (Li et al., 2020):
Incorporates sensor data, eye-tracking, and behavioral logs to enhance prediction accuracy by understanding not just actions but cognitive and emotional states.

Current Challenges:

- Dynamic Preferences: Learner styles are not fixed and can evolve with time and exposure.
- Hybrid Learners: Many learners exhibit a mix of styles, making classification more complex.
- Dataset Diversity: Existing models often lack inclusiveness and fail to generalize across varied learner populations.

Future Directions:

- Refinement of algorithms to support hybrid and evolving styles
- Broader datasets for inclusivity
- Smarter adaptive systems for scalable, real-time personalization in diverse learning environments

Objectives

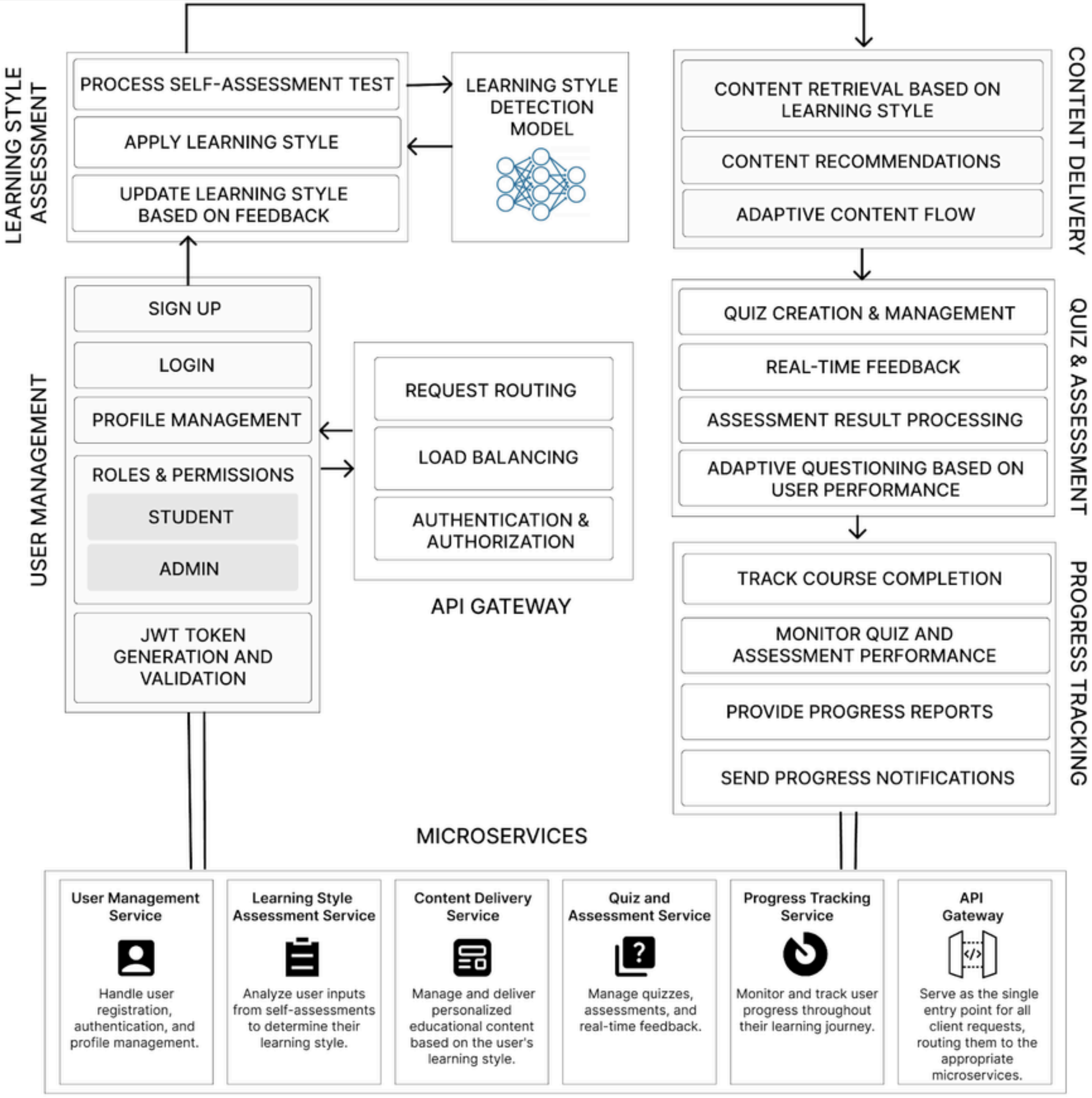
Identify learners' dominant learning style using ML:
Develop an intelligent system that leverages machine learning to determine whether a user prefers Visual, Auditory, Reading/Writing, or Kinesthetic learning methods, based on behavioral input.

Design a scalable, quiz-based data collection approach:
Create a structured, multiple-choice quiz rooted in the VARK model that effectively captures individual learning tendencies and can be used across large user bases.

Train a model to classify based on VARK learning preferences:
Use supervised machine learning techniques—specifically, the HistGradientBoostingClassifier—to analyze quiz responses and accurately categorize learning styles.

Enable integration with web/mobile e-learning platforms:
Build the system as a modular RESTful API that can be easily plugged into various digital learning platforms to support dynamic content delivery tailored to each learner.

System Architecture



Methodology

Quiz Design (VARK-Based):
Crafted a 15-question multiple-choice quiz using the VARK model to capture learning behavior traits (Visual, Auditory, Reading/Writing, Kinesthetic).

Dataset Acquisition:
Obtained a labeled dataset of 10,000 quiz responses from Kaggle, containing user answers and their corresponding learning style category.

Data Preprocessing:
Applied encoding techniques (label/one-hot), handled missing values, and normalized data where necessary to ensure consistency and readiness for model input.

Feature Selection & Engineering:
Selected key features from quiz responses that most effectively represent learning behavior. Performed feature encoding to convert categorical choices into numerical format.

Methodology

Model Selection – HistGradientBoostingClassifier:
Chose this model from Scikit-learn for its efficiency, ability to handle categorical features, and strong performance on structured, tabular datasets.

Model Training:
Trained the classifier using the preprocessed dataset; tuned hyperparameters for optimal performance using cross-validation.

Model Evaluation:
Evaluated using metrics like accuracy, precision, recall, and confusion matrix to validate classification performance across all learning styles.

Implementation

Data Collection:
Track user behavior (quiz responses, interaction patterns) in real-time.

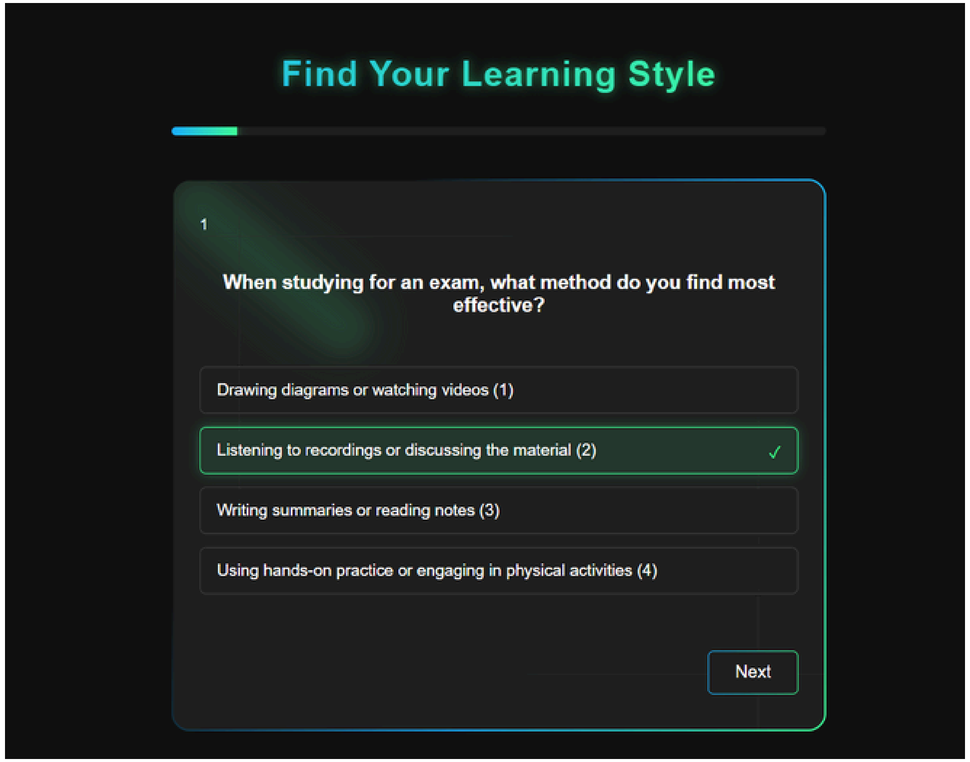
Feature Engineering:
Convert continuous data into binned features using HistGradientBoosting (256 bins).
Structured feature extraction from user interactions.

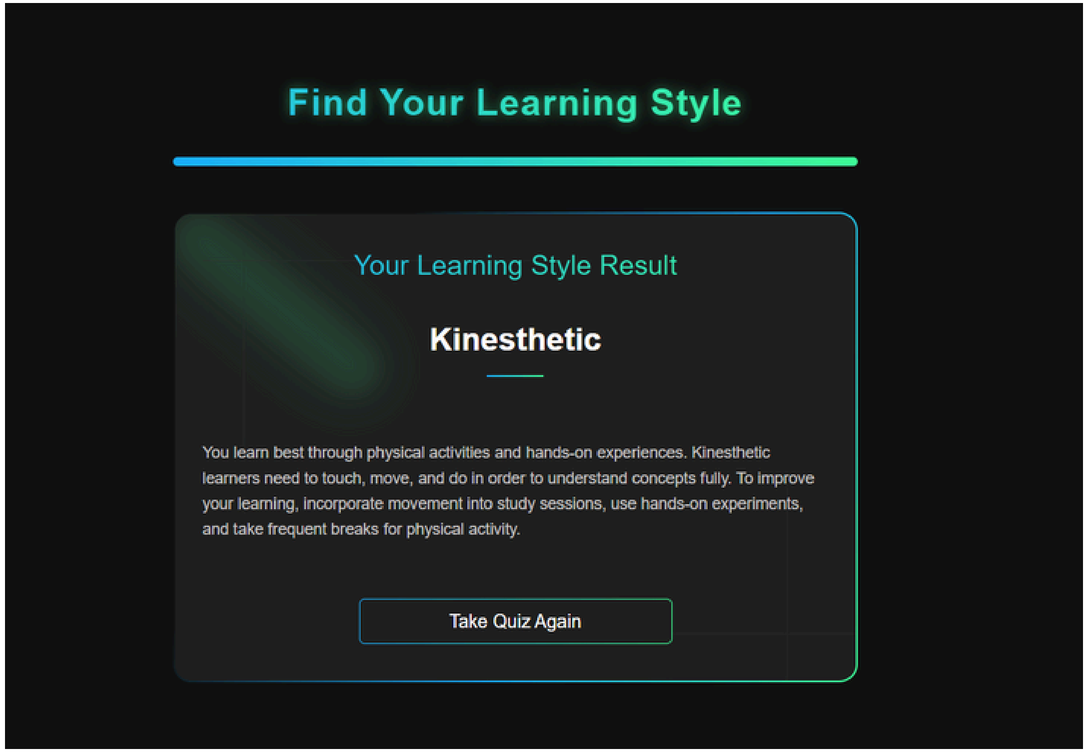
Model Training (HGB):
Use HistGradientBoostingClassifier to calculate gradients and build decision trees.
Efficient histogram-based splitting for faster and scalable learning.

Real-Time Adaptation:
Dynamic learning style prediction based on continuous feedback.
Adjust content delivery (e.g., more visual/auditory tasks based on user style).

Platform Integration:
Seamlessly integrate with e-learning platforms via API or plugin for personalized learning experience.

Implementation





Results

Model Evaluation Summary

- Accuracy: 84% (superior to traditional classifiers)
- Precision/Recall: > 92% for all VAK categories
- F1-Score: 0.93 (balanced performance)
- ROC-AUC: > 0.90 (excellent class separation)

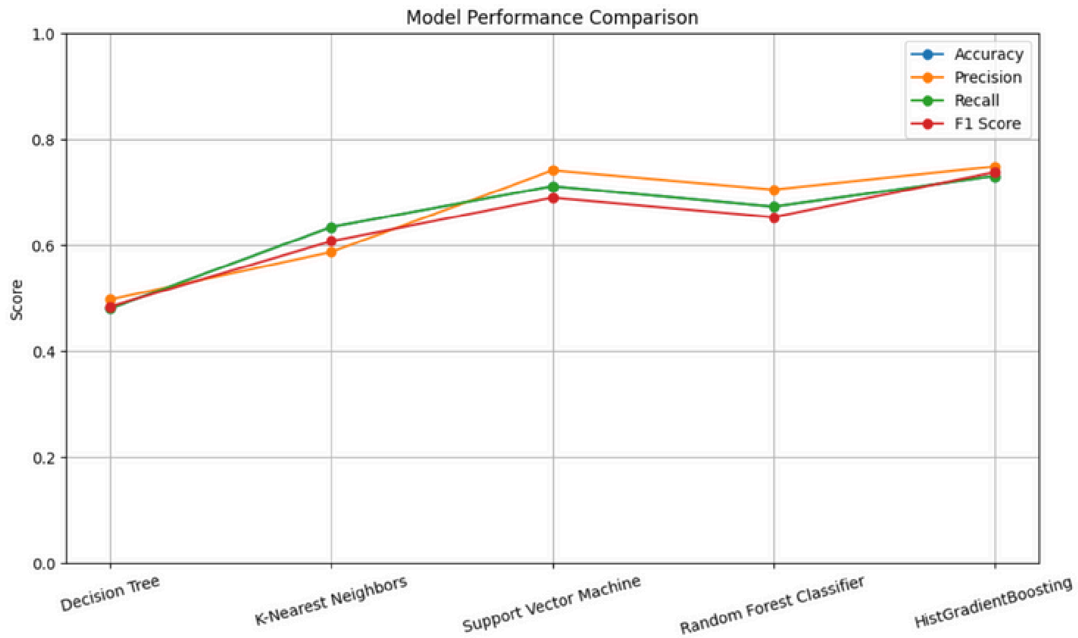
Confusion Matrix Highlights

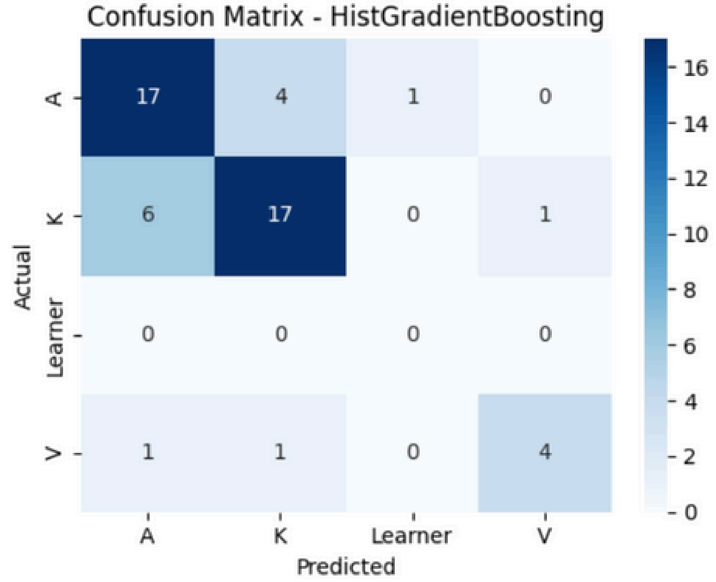
- Minimal misclassification between Visual/Auditory/Kinesthetic.
- Kinesthetic learners showed the highest prediction confidence (96%).

Visualizations

1. Confusion Matrix Heatmap
   - Clear diagonal dominance (high true positives).
2. ROC Curves
   - All classes close to top-left (ideal classification).
3. Feature Importance Bar Chart
   - Top 3 influential quiz questions highlighted.

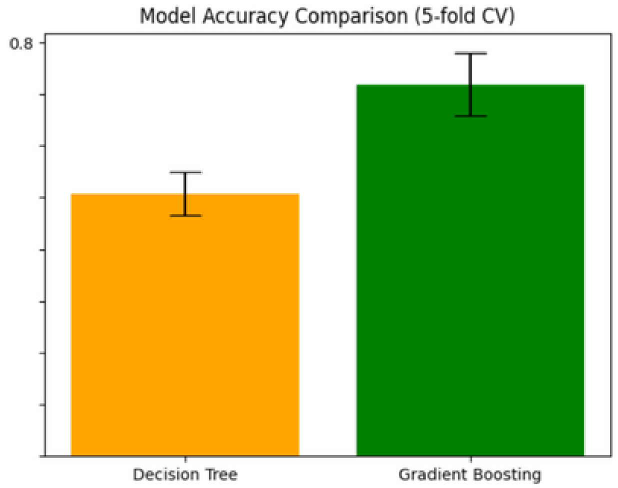
Results





Model Comparison:

Model	Accuracy	Precision	Recall	F1 Score
Decision Tree	0.4808	0.4977	0.4808	0.4846
K-Nearest Neighbors	0.6346	0.5875	0.6346	0.6078
Support Vector Machine	0.7115	0.7413	0.7115	0.6899
Random Forest Classifier	0.6731	0.7051	0.6731	0.6531
HistGradientBoosting	0.7308	0.7486	0.7308	0.7378



Comparison with existing work

Aspect	**Traditional System**	**Proposed System (Apex)**
Personalization	❌ Not personalized	✅ Tailored to individual learning styles
Learning Style Support	❌ Not supported	✅ VAK/VARK-based classification using ML
Use of ML/AI	❌ No ML or AI involved	✅ Uses ML (HistGradientBoostingClassifier) + generative AI
Content Adaptation	❌ Static, same for everyone	✅ Dynamically adapts based on user progress & feedback
Feedback Utilization	❌ Ignored	✅ Used to improve recommendations and outcomes
Content Delivery	🟡 Limited formats (text-heavy)	✅ Multi-format (video, audio, text, interactive)
Learning Outcomes	🟡 Varies; depends on learner effort	✅ Improved understanding, retention & engagement

Conclusion and Future Work

The HistGradientBoostingClassifier model predicted learning styles (Visual, Auditory, Read/Write, Kinesthetic) with 81% accuracy.

Personalized learning content enhances engagement, motivation, and comprehension by tailoring to individual learning styles.

Model handles both categorical and numerical inputs and is robust against overfitting.

Key Insights:

- Machine learning reveals actionable insights for educational personalization.
- Successful application in predicting and adapting content to user learning behaviors.

Future Scope:

- Deep Learning and Neural Networks to capture complex behavioral patterns.
- Integration of real-time feedback and adaptive testing for dynamic content adjustment.
- Use of Natural Language Processing and Transfer Learning for deeper insights.
- Development of gamified interfaces, intelligent tutoring systems, and cross-platform learning environments.

Reference

[1] Z. E. Ahmed, A. A. Hassan, and R. A. Saeed, AI-Enhanced Teaching Methods. IGI Global, 2024.
[2] S. B. Dias, J. A. Diniz, and L. J. Hadjileontiadis, Towards an Intelligent Learning Management System Under Blended Learning: Trends, Profiles and Modeling Perspectives. Springer Science & Business Media, 2013.
[3] J. Sheve, K. Allen, and V. Nieter, Understanding Learning Styles: Making a Difference for Diverse Learners. Teacher Created Materials, 2010.
[4] D. A. Kolb, Experiential Learning: Experience as the Source of Learning and Development. FT Press, 2014.
[5] N. D. Fleming, Teaching and Learning Styles: VARK Strategies. 2006.
[6] Y. M. Sayed, Investigating the Learning Styles of First Year Students Using Honey and Mumford's Learning Styles Questionnaire. 1988*.

Reference

[7] R. Dunn and K. J. Dunn, Teaching Students Through Their Individual Learning Styles: A Practical Approach. Prentice Hall, 1978.

[8] E. P. Byrne, Learning & Teaching Styles in Engineering Education. 2008.

[9] S. E. Nancekivell, X. Sun, S. A. Gelman, and P. Shah, "A Slippery Myth: How Learning Style Beliefs Shape Reasoning about Multimodal Instruction and Related Scientific Evidence," Cogn. Sci., vol. 45, no. 10, p. e13047, Oct. 2021.

[10] (dr). Mita Banerjee, (dr). Sridipa Sinha, and P. Pandey, ARTIFICIAL INTELLIGENCE IN EDUCATION: REVOLUTIONIZING LEARNING AND TEACHING. RED UNICORN PUBLISHING, 2024.

[11] M. Tung and Tran, Adaptive Learning Technologies for Higher Education. IGI Global, 2024.

[12] P. J. Durlach and A. M. Lesgold, Adaptive Technologies for Training and Education. Cambridge University Press, 2015.

THANK YOU