AI in Personalized Learning: SmartLearn AI

Student & Project Details

- **Student Name:** *Ashmita Luthra*
- ***Roll No:*** *iitrpr_ai_25010035*
- **Mentor Name:** Niranjan Deshpandey

- **Project Title:** *SmartLearn AI: Personalized Learning Path Recommendation System*

Abstract

This project presents **SmartLearn AI**, an AI-driven personalized learning system that adapts educational content based on student performance, engagement, and learning behavior. Traditional learning platforms use uniform content delivery, which fails to account for individual learning differences.

To address this, a supervised machine learning approach using a Random Forest classifier was implemented on a synthetic, unbiased dataset of 10,000 student records. The system predicts the most suitable next learning level (Easy, Medium, or Hard).

Experimental results show an overall accuracy of **62.95%**, reflecting realistic learning uncertainty while demonstrating effective personalization. The project highlights how AI can support adaptive learning in a transparent and explainable manner.

Introduction

Personalized learning is a critical requirement in modern education, as students differ in learning pace, engagement, and comprehension styles. Traditional rule-based or static learning systems often fail to adapt to these differences, leading to reduced effectiveness and engagement.

Recent advancements in AI and machine learning enable data-driven personalization by analyzing student behavior and performance patterns. This project focuses on building an AI-based system that dynamically recommends learning difficulty levels, aiming to enhance learning outcomes while maintaining ethical and explainable AI practices.

Related Work

Existing personalized learning systems often rely on rule-based heuristics or simple performance thresholds. More advanced systems use recommender systems and machine

learning models to adapt content based on user behavior. However, many approaches suffer from deterministic rules or lack explainability.

This project builds upon supervised learning techniques while explicitly addressing label leakage and incorporating controlled uncertainty to better reflect real-world educational environments.

Problem Definition

Objective

To design an AI system that recommends the next appropriate learning content level for students based on their learning behavior and performance.

AI Task

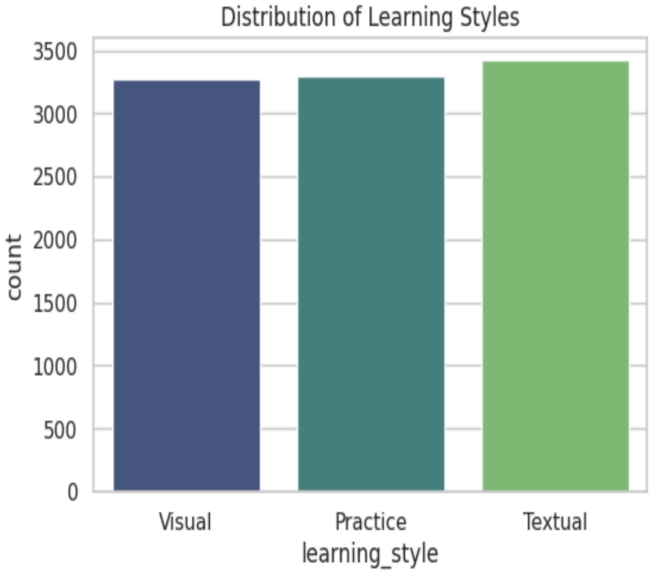
Multi-class classification (Easy, Medium, Hard).

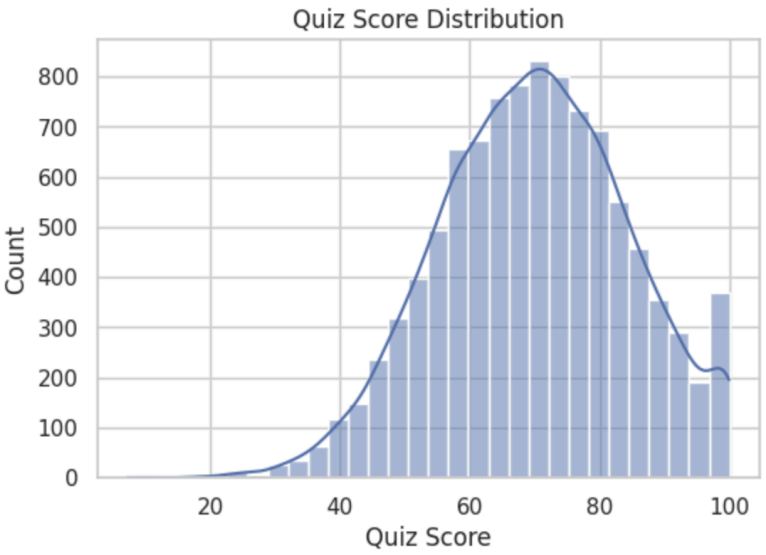
Constraints & Assumptions

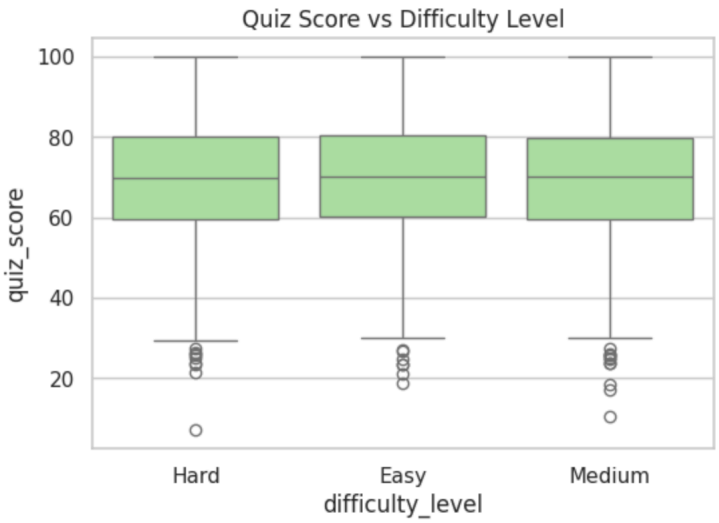
- Dataset is synthetic and unbiased.
- No personal or sensitive data is used.
- Learning progression is non-deterministic.
- The system supports educators rather than replacing human judgment.

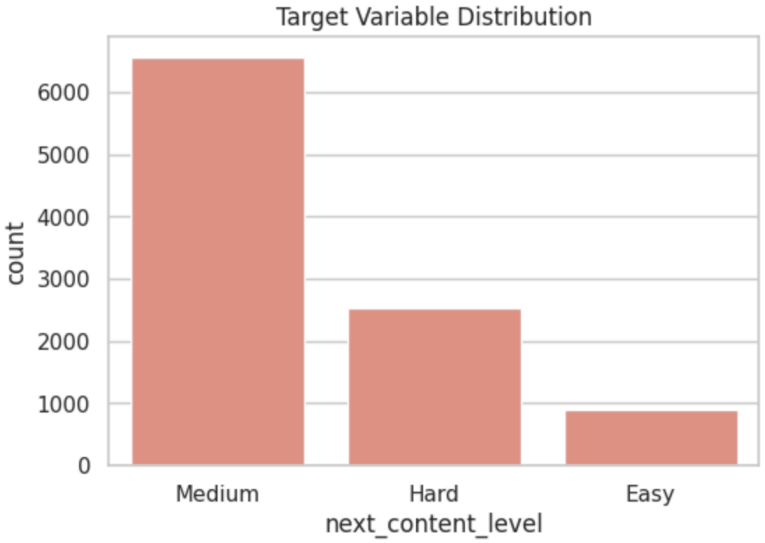
	student_id	topic	learning_style	difficulty_level	quiz_score	time_spent_minutes	attempts	engagement_score	next_content_level
0	1	Biology	Visual	Hard	78.3	35.1	4	51.7	Medium
1	2	Chemistry	Practice	Easy	51.8	49.3	1	45.7	Medium
2	3	Computer Science	Visual	Easy	65.7	24.2	2	44.1	Medium
3	4	Chemistry	Visual	Medium	69.3	30.8	3	46.9	Medium
4	5	Chemistry	Textual	Medium	69.6	45.8	4	50.5	Medium

Fig1: Dataset :personalized_learning_dataset_10k.csv









Methodology

System Overview

Student learning data is processed and fed into a supervised machine learning model that predicts the next learning level. The system includes preprocessing, model training, evaluation, and explainable recommendations.

Data Processing

- Synthetic dataset with 10,000 records.

- Features include quiz score, engagement score, attempts, time spent, learning style, and difficulty level.
- Categorical features encoded using Label Encoding.
- Numerical features standardized using StandardScaler.
- Stratified train-test split (80–20).

Model Design

- Algorithm: **Random Forest Classifier**
- Reason: Robustness, non-linearity handling, interpretability.
- Hyperparameter tuning applied to improve generalization.

Tools & Technologies

- Python
- Pandas, NumPy
- Scikit-learn
- Matplotlib, Seaborn
- Colab Notebook

Experiments and Results

Evaluation Metrics

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

Results

Overall Accuracy: 0.6295

	precision	recall	f1-score	support
Easy	0.69	0.30	0.42	583
Hard	0.73	0.61	0.66	411
Medium	0.59	0.83	0.69	1006
accuracy			0.63	2000
macro avg	0.67	0.58	0.59	2000
weighted avg	0.65	0.63	0.61	2000

Fig2: Classification Matrix

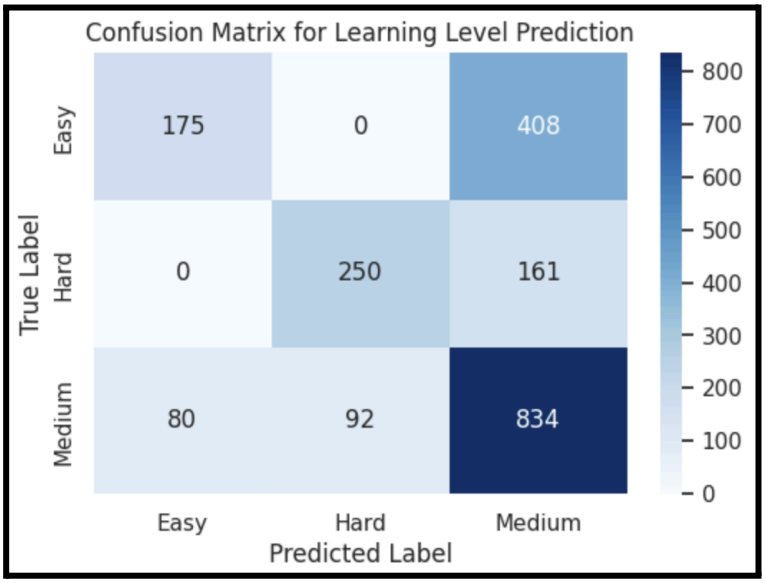


Fig3: Confusion Matrix for Learning Level Prediction

- The model performs best for **Medium-level learners**, reflecting realistic learning progression.
- Confusion matrix analysis shows most misclassifications occur between adjacent difficulty levels (Easy ↔ Medium), which is expected in educational contexts.
- Feature importance analysis highlights quiz score and engagement score as dominant predictors.

```
# Sample new student input

sample_student = pd.DataFrame([{
```

```
    "topic": "Computer Science",

    "learning_style": "Visual",

    "difficulty_level": "Medium",

    "quiz_score": 68,

    "time_spent_minutes": 55,

    "attempts": 2,

    "engagement_score": 72

}])
```

Output: `The system recommends **Easy** level content due to moderate quiz performance, good engagement.`

Conclusion and Future Work

This project successfully demonstrates a realistic AI-driven personalized learning system that adapts content recommendations based on multiple student attributes. The achieved accuracy reflects real-world uncertainty rather than overfitting or rule memorization. The system emphasizes ethical AI, explainability, and practical applicability.

Future Improvements

- Incorporating reinforcement learning for continuous adaptation.
- Using real-world educational data.
- Integrating large language models for personalized content generation.
- Extending recommendations to include learning format suggestions.

References & AI Usage Disclosure

- Dataset: Synthetic dataset generated for this project.
- Libraries: Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn.
- AI Assistance: Gemini & ChatGPT was used for guidance, debugging, and documentation support.