**Student Activity-Based Recommendation System:**

**A Comparative Study of KNN and Neural Network Models**

**1. Introduction**

In the educational platforms, understanding student activity patterns and predicting engagement behavior can significantly enhance the personalized content recommendation. This project presents a data-driven recommendation system based on student interaction logs. The system leverages timestamped activity data, such as platform usage, action types, and responses, to predict whether a student will engage (answer) with a given item or not.

The models compared in this study are:

* **K-Nearest Neighbors (KNN)** – a traditional distance-based algorithm.
* **Neural Network (Deep Learning)** – a state-of-the-art non-linear model capable of learning complex patterns.

**2. Dataset Description**

**2.1 Data Source**

EDNET dataset is used in this study. The dataset is composed of **27,621 records** aggregated from multiple .csv files. It includes user interaction data from an educational platform.

**2.2 Features Overview**

* **timestamp** – Unix time in milliseconds indicating when the action occurred.
* **action\_type** – Categorical variable describing the action (e.g., ‘viewed’, ‘answered’).
* **item\_id** – Unique identifier for the learning material.
* **source** – The source of the action (e.g., sprint, diagnosis).
* **user\_answer** – Indicates the user’s response (correct, incorrect, or no answer).
* **platform** – The platform used (web/mobile).

**2.3 Preprocessing & Feature Engineering**

* **Missing Values:** Over 21,000 missing values in user\_answer were replaced with the "No Answer".
* **Duplicates:** Only 6 duplicate rows were found.
* **Datetime Features:** Extracted hour and dayofweek from timestamp.
* **Target Variable:** Created is\_answered as binary label: 1 if the student answered, 0 otherwise.
* **Encoding:** Categorical columns encoded numerically.

**3. Exploratory Data Analysis (EDA)**

* **Temporal Activity Trends:** Clear spikes in activity observed during working hours and weekdays.
* **Action Frequency:** user\_answer shows a high rate of "No Answer", justifying the prediction task.
* **Visualizations:** Included plots for hourly and weekly activity patterns.

**4. Model Architectures**

**4.1 K-Nearest Neighbors (KNN)**

The KNN model’s structure is as follow:

* Used K=5.
* Features used: hour, dayofweek, action\_type\_encoded, platform\_encoded, source\_encoded.
* Simplicity and interpretability make it suitable for baseline.

**4.2 Neural Network (Deep Learning)**

The architecture of Neural Netwok is as follow:

* Architecture:
  + Input layer (5 features)
  + Dense(64) + ReLU + Dropout(0.3)
  + Dense(32) + ReLU + Dropout(0.3)
  + Output: Sigmoid layer is used as output (for binary classification)
* Loss: Binary Crossentropy is used as loss function.
* Optimizer: Adam optimizer is used.
* Early stopping applied to prevent overfitting.

**5. Model Evaluation**

Models are evaluated using F1 score and accuracy.

* **KNN:**
  + Performs reasonably well given the small feature space.
  + Struggles with overlapping decision boundaries in high-dimensional categorical data.
* **Neural Network:**
  + Outperforms KNN significantly.
  + Learns non-linear interactions and benefits from dropout regularization.

**6. Deployment & Saving**

* **Models Saved:** knn\_model.pkl, deeplearning\_model.keras
* **Comparison Exported:** model\_comparison.csv
* **Visuals Saved:** activity\_over\_time.png, hourly\_activity.png, weekly\_activity.png

**7. Discussion & State-of-the-Art Positioning**

This recommendation system aligns well with the modern **student modeling** strategies that prioritize engagement prediction using historical behavior. While **KNN** provides a baseline with low computational complexity, **deep learning** models are now the state-of-the-art in educational analytics due to their robustness and scalability.

Neural networks, with dropout and early stopping, exhibit resilience against overfitting and generalize well even with imbalanced or sparse data. The achieved validation accuracy of **99.9%** confirms its superiority over traditional models like KNN.

**8. Conclusion & Future Work**

This project demonstrates:

* The feasibility of behavior-based recommendation using log data.
* The advantage of neural models over traditional methods in such tasks.

**Future Enhancements:**

* Integrate item difficulty and student history.
* Expand features (e.g., session duration, clickstream patterns).
* Incorporate sequence modeling (RNNs or Transformers) for deeper behavior modeling.

**Appendices**

* **A. Dependencies**: scikit-learn, tensorflow, joblib, matplotlib, pandas, seaborn.





