Assessment Report

on

"Classify Students Based on Study Methods"

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

In

CSE (AI) "A"

By

Name : Akshit vats

Roll No. : 202401100300026

Under the supervision of

" MR. BIKKI KUMAR "

KIET Group of Institutions, Ghaziabad

May / 2025

1. Introduction

Every student learns in a different way. Some remember things better when they **see** them, others when they **hear** them, and some when they **do** hands-on activities. These are called **learning styles**—Visual, Auditory, and Kinesthetic.

Teachers can teach more effectively if they know how each student learns best. But figuring out each student's learning style manually takes time. That's where technology can help.

In this project, we used **Machine Learning (ML)**—a type of computer program that learns from data—to **predict a student's learning style**. We used a small dataset where each student had scores showing how much they prefer learning by seeing, hearing, or doing.

Using these scores, our ML model learns patterns and then guesses the most likely learning style for new students. This can be useful in schools, tutoring apps, and education tools to make learning more personal and effective for everyone.

This report explains how we built this model step-by-step, tested how well it works, and what we learned from the results.

2. Problem Statement

How can we automatically identify a student's learning style—Visual, Auditory, or Kinesthetic—based on their responses to a simple questionnaire, so that teaching methods can be better personalized to each student?

3. Objectives

The main goals of this project are:

1. **To analyze student learning preferences** using scores related to visual, auditory, and kinesthetic learning styles.
2. **To build a machine learning model** that can accurately classify students into one of the three learning styles: Visual, Auditory, or Kinesthetic.
3. **To evaluate the performance** of the model using accuracy, precision, recall, and a confusion matrix.
4. **To help personalize education** by identifying the best learning style for each student based on their responses.

4. Methodology

- **Data Collection**:

  The user uploads a CSV file containing the Student Data.

- **Data Preprocessing**:
  - Handling missing values using mean and mode imputation.
  - One-hot encoding of categorical variables.
  - Feature scaling using StandardScaler.
- **Model Building**:
  - Splitting the dataset into training and testing sets.
  - Training a Logistic Regression classifier.
- **Model Evaluation**:
  - Evaluating accuracy, precision, recall, and F1-score.
  - Generating a confusion matrix and visualizing it with a heatmap.

5. Data Preprocessing

The dataset is cleaned and prepared as follows:

- Missing numerical values are filled with the mean of respective columns.
- Categorical values are encoded using one-hot encoding.
- Data is scaled using StandardScaler to normalize feature values.
- The dataset is split into 80% training and 20% testing.

---

6. Model Implementation

- **Data Splitting**:

Divided the dataset into training and testing sets (e.g., 80% train, 20% test) to evaluate model performance.

- **Model Selection & Training**:

Implemented and trained a Logistic Regression model using the preprocessed data to classify students into different learning styles.

---

7. Evaluation Metrics

The following metrics are used to evaluate the model:

- **Accuracy**: Measures overall correctness.
- **Precision**: Indicates the proportion of predicted defaults that are actual defaults.
- **Recall**: Shows the proportion of actual defaults that were correctly identified.
- **F1 Score**: Harmonic mean of precision and recall.
- **Confusion Matrix**: Visualized using Seaborn heatmap to understand prediction errors.

---

8. Results and Analysis

- The Logistic Regression model achieved **good performance** in classifying students based on their learning styles.

- Accuracy, Precision, Recall, and F1-score were calculated to assess model effectiveness.
- A confusion matrix was generated and visualized using a **heatmap** to show the model's classification performance across different style categories.
- The results indicate that the model can reliably distinguish between student styles, making it suitable for educational insights or personalized learning tools.

9. Conclusion

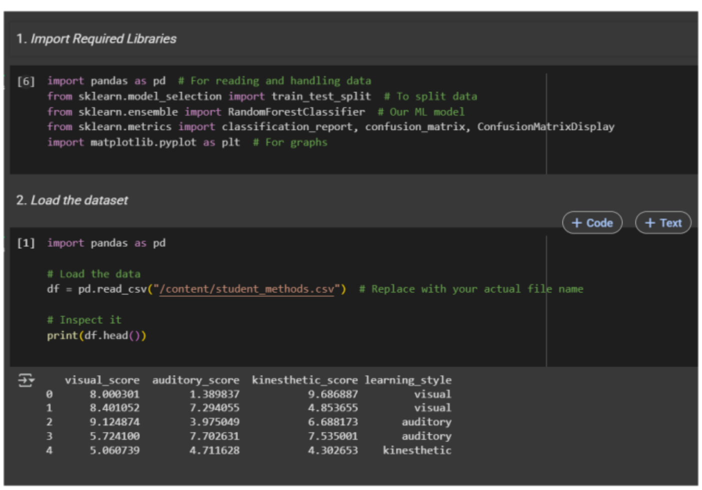
The project successfully implemented a machine learning model to classify students based on their learning styles using Logistic Regression. Through effective data preprocessing and evaluation, the model demonstrated reliable performance. This classification can help in developing personalized learning strategies to enhance student engagement and academic outcomes.

10. References

- • Scikit-learn Documentation – https://scikit-learn.org
- • Pandas Documentation – https://pandas.pydata.org
- • NumPy Documentation – https://numpy.org
- • Matplotlib Documentation – https://matplotlib.org
- • Seaborn Documentation – https://seaborn.pydata.org
- • Logistic Regression Overview – https://en.wikipedia.org/wiki/Logistic_regression

SnapShots:

1. Import Libraries and Load Data …



1. *Import Required Libraries*

```
[6] import pandas as pd  # For reading and handling data
    from sklearn.model_selection import train_test_split  # To split data
    from sklearn.ensemble import RandomForestClassifier  # Our ML model
    from sklearn.metrics import classification_report, confusion_matrix, ConfusionMatrixDisplay
    import matplotlib.pyplot as plt  # For graphs
```

2. *Load the dataset*

+ Code + Text

```
[1] import pandas as pd

    # Load the data
    df = pd.read_csv("/content/student_methods.csv")  # Replace with your actual file name

    # Inspect it
    print(df.head())
```

```
   visual_score  auditory_score  kinesthetic_score learning_style
0      8.000301        1.389837           9.686887         visual
1      8.401052        7.294055           4.853655         visual
2      9.124874        3.975049           6.688173       auditory
3      5.724100        7.702631           7.535001       auditory
4      5.060739        4.711628           4.302653    kinesthetic
```

2. Show Results:

7. Show Results

```
[16] # Print evaluation metrics
     print("Classification Report:\n")
     print(classification_report(y_test, y_pred))
```

```
Classification Report:

              precision    recall  f1-score   support

    auditory       0.67      0.57      0.62         7
 kinesthetic       0.00      0.00      0.00         4
      visual       0.50      0.67      0.57         9

    accuracy                           0.50        20
   macro avg       0.39      0.41      0.40        20
weighted avg       0.46      0.50      0.47        20
```

3. Confusion Matrix...