****

**Assessment Report**

on

**“Student Club Participation Prediction”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AIML)**

By

Name : Abik Thakur

Roll Number : 202401100400010

Section: ‘A’

**Under the supervision of**

“BIKKI GUPTA”

**KIET Group of Institutions, Ghaziabad**

**April, 2025**

1. **Introduction**

**In today's educational landscape, understanding student engagement is crucial to improving both academic performance and overall student experience. This report explores a dataset focusing on students' interest levels, free hours per week, and their participation in clubs. The objective is to identify patterns in student engagement, including the relationship between club participation and students' interest levels.**

**2. Problem Statement**

Student engagement plays a significant role in enhancing academic performance and extracurricular involvement. This report seeks to address the following questions:

1. **How does participation in clubs impact student interest levels?**
2. **What is the average interest level of students with varying free hours per week?**
3. **Is there a noticeable difference in engagement between students who are part of clubs versus those who are not?**

By analyzing these factors, we aim to provide insights that could help institutions make informed decisions on fostering student engagement.

### ****3. Objectives****

● Preprocess the agricultural dataset for machine learning.  
● Train a Logistic Regression model to classify crop yield into categories.  
● Evaluate model performance using standard classification metrics.  
● Visualize the confusion matrix using a heatmap for better interpretability.



### ****4. Methodology****

The dataset consists of the following columns:

* **Interest Level**: Numerical representation of student interest (scale of 1 to 10).
* **Free Hours per Week**: Number of hours a student has free in a week.
* **Club Participation**: Whether or not a student participates in a club ("yes" or "no").

### ****5. Data Preprocessing****

The dataset was processed using the following steps:

* **Loading Data**: The dataset was uploaded and loaded into a pandas DataFrame for analysis.
* **Handling Categorical Data**: The club\_participation column, which had "yes" and "no" values, was converted to boolean values (True for "yes" and False for "no") for easier analysis.
* **Grouping Data**: Data was grouped by club\_participation to calculate the average interest\_level for each group, i.e., club members versus non-club members.
* **Visualization**: A bar chart was plotted to visually compare the average interest level between the two groups: those in clubs and those not in clubs.

**6. Model Implementation**

**Logistic Regression is used as it is a strong baseline classifier for multiclass problems. The model is trained on the processed dataset and used to predict the yield\_category.**

### ****7. Evaluation Metrics****

The evaluation of this dataset can be approached using the following metrics:

1. **Interest Level Difference**: A comparison of the average interest level between club participants and non-participants. This provides insights into the impact of extracurricular engagement on student interest.
2. **Data Consistency**: Ensuring that the dataset does not have missing or inconsistent values.
3. **Visual Insights**: The clarity of visualizations (e.g., bar graphs) that clearly highlight patterns in student engagement.
4. **Statistical Analysis**: Using averages and percentages to evaluate engagement differences effectively.

In this case, the primary evaluation focuses on whether participating in a club increases students' interest levels compared to those who are not part of any club.

### ****8. Results and Analysis****

● The model showed acceptable performance on the test set.  
● The confusion matrix illustrated how well the model differentiated between yield categories.  
● Precision and recall metrics revealed strengths and weaknesses in predicting specific classes.

**9. Conclusion**

**Summary of Findings:**

* **Average Interest Level**: The average interest level for **club members** was found to be higher compared to those who are **not in clubs**.
  + **Club Members**: The average interest level is higher due to the additional engagement that club activities offer.
  + **Non-Club Members**: The average interest level was comparatively lower, suggesting that club participation has a positive impact on student engagement.

### ****10. References****

● scikit-learn documentation  
● pandas documentation  
● Seaborn visualization library  
● Research articles on crop yield prediction and agricultural analytics



