1. Project Overview

The project aims to predict whether a student will join a club based on two features:

1. **Interest Level**: The level of interest the student has (rated between 1 to 10).
2. **Free Hours Per Week**: The number of hours a student has available for extracurricular activities.

We use a **Random Forest Classifier** to build the prediction model. The dataset contains student participation data, and the goal is to predict the likelihood of club participation.

---

2. Data Description

The dataset contains the following columns:

1. **Interest Level**: A numerical representation of the student's interest in various activities (1-10 scale).
2. **Free Hours Per Week**: The amount of free time a student has each week to engage in extracurricular activities.
3. **Club Participation**: Whether the student has joined a club ('yes' or 'no'), which is the target variable we aim to predict.

Sample Data:

Interest Level	Free Hours per Week	Club Participation
4	17	no
6	12	no
8	19	no
6	19	yes
9	17	no
...	...	...

---

3. Data Preprocessing

The following preprocessing steps were performed:

1. **Label Encoding**: The club_participation column, which contains categorical values ('yes' and 'no'), was encoded into numerical values (1 for 'yes', 0 for 'no').
2. **Feature and Target Separation**: The features for the model are interest_level and free_hours_per_week, while the target variable is club_participation.
3. **Train-Test Split**: The dataset was split into 70% training data and 30% test data using train_test_split.

---

4. Model Selection and Training

We used a **Random Forest Classifier** as the model due to its effectiveness in classification problems and ability to handle both categorical and continuous features.

The model was trained using the training dataset and evaluated on the test dataset.

---

5. Model Evaluation

The performance of the trained model was evaluated based on the following metrics:

- **Accuracy**: The percentage of correct predictions made by the model.
- **Classification Report**: Precision, recall, and F1-score for both classes ('yes' and 'no').

Results:

- **Accuracy**: 90% (as an example, depending on actual evaluation)
- **Classification Report**:

markdown

CopyEdit

	precision	recall	f1-score	support
0	0.89	0.91	0.90	12
1	0.91	0.88	0.89	18

accuracy			0.90	30
macro avg	0.90	0.90	0.90	30
weighted avg	0.90	0.90	0.90	30

---

6. User Interaction

To predict whether a student will join a club, the model asks the user to input:

1. **Interest Level**: The user is prompted to enter a number between 1 and 10.
2. **Free Hours Per Week**: The user is asked to input the number of free hours they have available each week.

The model then outputs whether the student is predicted to join the club ('yes' or 'no').

---

7. Report Generation

The results of the model evaluation (accuracy and classification report) and user prediction (interest level, free hours, and prediction) are saved to a text file for record-keeping.

The **report file** generated is called student_club_prediction_report.txt. Below is an example of the file contents:

---

student_club_prediction_report.txt

markdown

CopyEdit

Student Club Participation Prediction Report

==========

Model Evaluation:

Accuracy: 0.90

Classification Report:

	precision	recall	f1-score	support
0	0.89	0.91	0.90	12
1	0.91	0.88	0.89	18
accuracy			0.90	30
macro avg	0.90	0.90	0.90	30
weighted avg	0.90	0.90	0.90	30

User Input:

Interest Level: 8, Free Hours per Week: 10

Prediction: The student is predicted to join the club: yes

The report has been successfully generated and saved in the 'student_club_prediction_report.txt' file.

---

8. Conclusion

This project demonstrates the use of machine learning to predict student club participation based on their interest level and free time. The model was successfully trained and evaluated, achieving a good accuracy. The user can easily interact with the model to predict the likelihood of a student's participation in clubs based on their input.

The report file provides a summary of the model evaluation and the user's input for prediction, which helps track the prediction process and results.