# Model Usage Instructions

## Introduction

Welcome to our Bacterial Culture Medium Prediction Models. These 45 models are built using the XGBoost machine learning algorithm, designed to help researchers and laboratory technicians accurately predict the best culture medium for different types of bacteria. The models have been extensively trained on a wide range of data, providing high-accuracy predictions to optimize laboratory workflows, saving time and resources.

## Preparation

### System Requirements

Python 3.7 or higher

Pandas library

Pickle library

iLearnPlus software installed, for processing 16S rRNA sequence data

Make sure your system has the above software and libraries installed. Installation guides for iLearnPlus can usually be found on its official website or release page.

### Data Preparation

**1.Download 16S rRNA Sequences:** Download 16S rRNA sequence data from the database you are using.

**2.Generate Feature Data with iLearnPlus:**

Open the iLearnPlus software.

Use the RNA's kmer function to convert the downloaded 16S rRNA sequence information into feature data.

Save the generated feature data as a CSV file, ensuring the first column is set as the data index.

### Model Preparation

Download Models: Download and save the 45 trained models into a specified folder. Ensure each model file has the .pkl extension.

## Usage Steps

### Modify Script Paths:

Adjust the model\_dir variable to the path of your model folder, according to where your files are stored.

Change the data\_path variable to the path of your new dataset file (i.e., the CSV file generated with iLearnPlus software).

Modify the result\_path variable to where you wish to save the CSV file with prediction results.

### Execute the Script:

Run the script. It will automatically load each model, use them to make predictions on the new data, and compile the prediction results into a DataFrame.

### View Prediction Results:

The prediction results will be saved to the path you specified, with the filename result\_use.csv. You can view these results with any text editor or the Pandas library.

## Points to Note

Ensure that when processing sequence data with iLearnPlus software, you choose the same feature extraction method as was used during model training.

The predictive capacity of the models is limited by the quality and range of their training data. Performance may vary when applied to new data that significantly differs from the training data.