

Breaking Bug - Machine Learning Repository

Table of Content

- [Introduction](#)
- [Pre-requisites](#)
- [How to get started?](#)
- [Setting up the project](#)
- [Project Information](#)
- [Credits](#)

Introduction

This repository contains the backend code for the Breaking Bug event. The event is organized by IEEE Computer Society, Manipal University Jaipur.

Breaking Bug is an electrifying virtual showdown for tech enthusiasts and coding maestros! An exciting and challenging event where participants step into the shoes of skilled developers and problem-solvers! In this unique competition, their mission is to identify and fix bugs in a GitHub repository across three diverse domains: Frontend, Backend, and Machine Learning (ML).

Pre-requisites

- Python
- Pandas
- NumPy
- Matplotlib
- Scikit-learn
- XGBoost

How to get started?

- Download the code and dataset from the provided links.
- Set up your development environment with the pre-requisites listed above.
- Follow the steps mentioned in the [Setting up the project](#) section.
- Use any platform or IDE to fix the bug, test, and train your model.
- After making the changes, push your code to your cloned repository.
- Submit the link to your updated repository or results through the specified form.

Setting up the project

- Fork the repository by clicking on the Fork button on the top right corner of the page.
- Clone the repository in your local system.

```
git clone https://github.com/<your-github-username>/BreakingBug-ML.git
```

- Navigate to the project directory.

```
cd BugBytes-ML
```

- Download the dataset `dataset.csv`.
- Open the `bugged_file.py` and review the code to identify and fix the bugs.
- Test and train your model using the dataset.
- After making the necessary changes, run the script to ensure everything works correctly.
- Push your changes to your cloned repository.
- Submit the link to your updated repository or results through the form link provided.

Project Information

Here's a revised summary focusing on the ML-related details:

Points Distribution

The maximum attainable points for this project are 1000. The points are distributed as follows:

Difficulty Level	Points
Very easy	20
Easy	30
Medium	40
Hard	75
Easter egg	100
Total	1000

Here are the columns from the dataset, with their descriptions:

Dataset Columns

- **id**: Unique ID
- **age**: Age in years

- **sex:** Gender
- **dataset:** Location of data collection
- **cp:** Chest pain type
- **trestbps:** Resting blood pressure
- **chol:** Cholesterol measure
- **fbs:** Fasting blood sugar
- **restecg:** ECG observation at resting condition
- **thalch:** Maximum heart rate achieved
- **exang:** Exercise induced angina
- **oldpeak:** ST depression induced by exercise relative to rest
- **slope:** The slope of the peak exercise ST segment
- **ca:** Number of major vessels (0-3) colored by fluoroscopy
- **thal:** Thalassemia
- **num:** Target [0 = no heart disease; 1, 2, 3, 4 = stages of heart disease]

Here's the updated table without the "Best Hyperparameters" column:

Model Performance

Model	Cross-Validation Accuracy	Test Accuracy
Logistic Regression	0.5115	0.5109
Gradient Boosting	0.6396	0.5978
KNeighbors Classifier	0.5767	0.5870
Decision Tree Classifier	0.5840	0.5761
AdaBoost Classifier	0.6058	0.5978
Random Forest	0.6288	0.6739
XGBoost Classifier	0.6263	0.6413
Support Vector Machine	0.5877	0.5870
Naive Bayes Classifier	0.5780	0.5435

Best Model: XGBoost Classifier

Best Model Cross-Validation Accuracy: 0.6263

Best Model Test Accuracy: 0.6413

Made by IEEE Computer Society- Manipal University Jaipur