# Identifying Bug Types & Severity in Open-Source Code

Arhaan Bahadur
2022093

Ishir Bhardwaj
2022233

Manit Kaushik
2022277

Pranav Gupta
2022364

Raghav Wadhwa
2022385

## 1. Project Proposal

First submission for the group project part of the CSE363 Machine Learning Course taught by Prof. Jainendra Shukla.

### 1.1. Motivation

As engineering students, we've faced debugging challenges while creating projects or learning new technologies. Recognizing the need for a more efficient solution, we aim to develop a machine learning-based approach to automatically identify bug type and severity, streamlining the debugging process.

### 1.2. Related Work

Recent work done on this subject matter -

[1] A. Baarah, A. Al-oqaily, Z. Salah, M. Sallam, & M. Al-qaisy. (2019), Machine Learning Approaches for Predicting the Severity Level of Software Bug Reports in Closed Source Projects, IJACSA.

This paper uses different Machine Learning approaches for prediciting severity of bugs in the JIRA bug tracking system of INTIX Company.

[2] Tan, Y., Xu, S., Wang, Z., Zhang, T., Xu, Z., & Luo, X. (2020). Bug Severity Prediction Using Question-and-Answer Pairs from Stack Overflow. Journal of Systems and Software, 110567.

This paper proposes a novel approach to predict the severity of bug reports using question-and-answer pairs from Stack Overflow and machine learning algorithms.

[3] This paper proposes a taxonomy of bug types and an automated classification model to classify reported bugs in popular projects across three ecosystems.

Catolino, G., Palomba, F., Zaidman, A.,& Ferrucci, F. (2019). Not All Bugs Are the Same:Understanding, Characterizing, and Classifying Bug Types. Journal of Systems and Software.

### 1.3. Timeline

The proposed plan of action for the project is outlined below:

1. **Literature Review and Goal Setting**: Review of existing research and define project objectives.
2. **Dataset Acquisition and Preprocessing**: Identify suitable datasets for bug classes and severity.
3. **Feature Engineering**: Extract relevant features for classification and regression tasks.
4. **Baseline Establishment**: Develop baseline models using pre-existing practices to establish a benchmark.
5. **Model Development**: Design and implement separate models for bug identification and severity prediction.
6. **Model Training and Fine-Tuning**: Train and refine the models using the prepared datasets.
7. **Performance Evaluation**: Assess the performance of the models using various evaluation metrics.
8. **Project Finalization**: Final presentation and project paper creation.

### 1.4. Individual Tasks

Each member will be assigned a specific Machine Learning paradigm like regression, CNN etc. which will be part of either of the 2 models (severity calculation and bug classification). Given here are the tasks and the people assigned to it:

- Literature Review: Arhaan, Pranav, Raghav
- Setting Baselines: Arhaan, Manit
- Feature Engineering: Ishir, Raghav
- Bug Type Classification Model: Ishir, Arhaan
- Bug Severity Calculation Model: Manit, Pranav
- Model Evaluation: Manit, Pranav
- Project Presentation and Paper: Ishir, Raghav

### 1.5. Final Outcome

We aim to contribute to the existing body of research in this area and provide a valuable tool for software developers and maintainers. By the end of this project, we aim to develop a robust and accurate machine-learning model for bug type and severity identification and how it fairs against existing techniques.