# Software Defect Prediction Using Ensemble Techniques and XAI

| | |
|---|---|
| Student Name | D.M Sandali P Dissanayake |
| IIT ID | 20211576 |
| RGU ID | 2121985 |
| Submission | Final Report Submission |
| Coursework Title | CMM799 – MSc Project – BA |
| Supervisor Name | Mr. Guhanathan Poravi |

**CONSENT**

I agree ■

I do not agree ☐

That the University shall be entitled to use any results, materials or other outcomes arising from my project work for the purposes of non-commercial teaching and research, including collaboration.

**DECLARATION**

**I confirm:**

- **That the work contained in this document has been composed solely by myself and that I have not made use of any unauthorised assistance.**
- **That the work has not been accepted in any previous application for a degree.**
- **All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.**

**Abstract**

The use of machine learning (ML) techniques for predicting software defects is the main goal of this study. By addressing challenges including redundancy, correlation, and feature irrelevance, ensemble learning assists ML models perform better. An in-depth analysis of Software Defect Prediction (SDP) employing ensemble approaches and machine learning techniques with explainable artificial intelligence (XAI) is presented in this study. It investigates the use of ensemble approaches, such as stacking, to combine the capabilities of various base models, including Random Forest, Naïve Bayes, SVM, Logistic Regression, XGBoost, AdaBoost, and Decision Trees, in order to improve the accuracy of defect prediction models while employing a variety of preprocessing techniques including SMOTE. Stacking ensemble proved to be the best model while achieving 80% of predictive accuracy. Among individual model training, Random Forest performed the best achieving 79% of predictive accuracy. Logistic Regression achieved the lowest predictive accuracy with 65%.

It underlines the need of XAI techniques in Software Defect Prediction in addition to model creation. By utilizing SHAP values and LIME, it offers insightful explanations for model predictions and useful insights into the elements causing software defects. By bringing transparency to complicated black-box models, these XAI approaches assist stakeholders and software developers better understand and utilize the defect prediction process.

Overall, this study makes a significant contribution to the field of software defect prediction by emphasizing the role of ensemble approaches and XAI in improving predictive insights and providing useful information for software development. Combining modern machine learning techniques with understandable explanations creates new opportunities for accurate and useful defect prediction, which ultimately helps software development practices and the production of high-quality software products.