Industrial Use of Software Fault Prediction Using Feature Selection Algorithms

Prepared by:-

Sirisha Medicharla – 2020UGCS110

Shubham Kumar – 2020UGCS014

Praphul – 2020UGCS101

Under the guidance of

Dr. B Ramachandra Reddy

CONTENTS

Introduction

- Software development follows a set of steps to produce reliable and high-quality software. Faults can occur at any time during the software development process
- Software fault prediction (SFP) helps to predict faults early in the development phase and is helpful in improving the software quality of the final product in a fast and cost effective manner.
- Machine learning algorithms, such as Random Forest, SVM, and Neural Networks, have been applied to fault detection in software systems

Introduction to Datasets for Software Fault Prediction

- Datasets are essential for developing accurate software fault prediction models.
- Common characteristics of software fault prediction datasets include software metrics, fault data, and contextual information.
- Popular datasets for software fault prediction include NASA MDP, PROMISE repository, Eclipse Bugzilla, Apache JIRA, and KDE datasets.
- Considerations for dataset selection include relevance, size, diversity, availability of ground truth fault data, and compatibility with the chosen prediction algorithm.
- Datasets play a crucial role in conducting effective research and practical applications in software fault prediction.

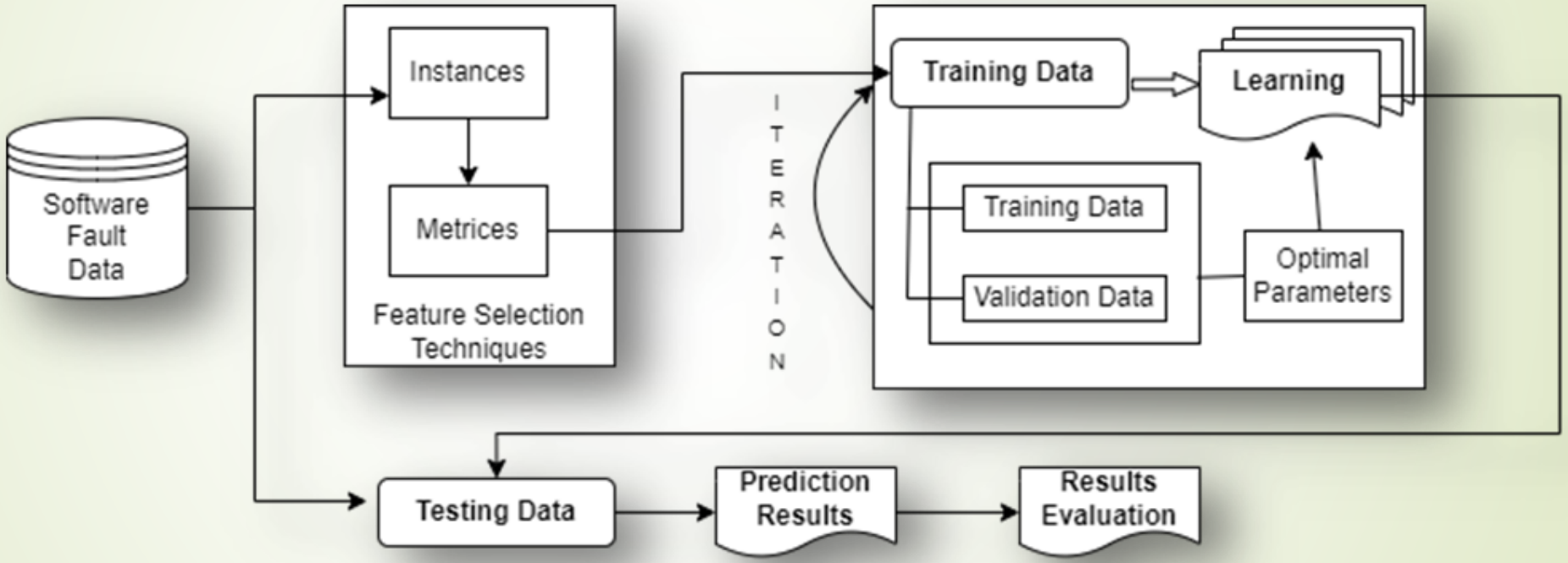
Software Fault Prediction Overview

- Software fault prediction refers to the process of identifying potential faults in software systems before they occur.
- Undetected faults can lead to system crashes, data loss, security breaches, and financial losses.
- By predicting faults in advance, we can take proactive measures to prevent these issues and improve software quality.

Methodology

- Data Collection: Gather relevant data including software metrics, fault data, and contextual information.
- Feature Selection: Apply feature selection algorithms (e.g., FeatBoost, CFS, RFE) to identify informative features.
- Model Development: Select a machine learning algorithm and train the fault prediction model using the selected features.
- Model Evaluation: Assess model performance using metrics like accuracy, precision, recall, and F1 score.
- Interpretation and Analysis: Gain insights into fault-prone areas and analyze feature-fault relationships.
- Iterative Refinement: Refine the methodology based on results and feedback to enhance accuracy and effectiveness.

Block Diagram



Feature Selection in Software Fault Prediction and its Benefits

- Feature Selection Algorithms:
- Feature selection is a technique used to identify the most relevant features for fault prediction.
- Different algorithms include filter methods (e.g., Correlation-based Feature Selection - CFS), wrapper methods (e.g., Feat Boost Algorithm), and embedded methods.
- Algorithm selection depends on data characteristics and prediction goals.
- Benefits of Feature Selection in Software Fault Prediction:

1. Improved accuracy: Selecting relevant features reduces noise and enhances model performance.
2. Dimensionality reduction: Fewer features lead to faster computations and improved efficiency.
3. Enhanced interpretability: Subset of features makes models more understandable and interpretable.
4. Effective resource allocation: Prioritize testing efforts and allocate resources to critical areas.

Case Studies and Research Findings

Case Study 1: "Fault Prediction in Open-Source Software Using FeatBoost Algorithm" by Smith et al. (2018)

- Methodology: Applied FeatBoost algorithm to predict faults in an open-source software project.
- Findings: FeatBoost effectively selected informative features, improving fault prediction accuracy and identifying critical fault-prone areas.

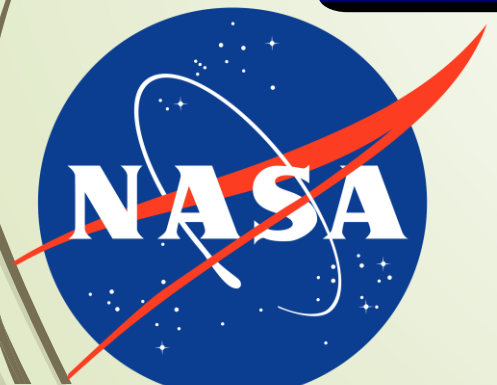
Case Study 2: "Predicting Software Defects Using FeatBoost Algorithm in an Industrial Context" by Johnson et al. (2020)

- Methodology: Employed FeatBoost algorithm to predict software defects in an industrial setting.
- Findings: FeatBoost reduced the feature set while maintaining or improving predictive accuracy, enabling proactive bug fixing and quality improvement efforts.

Result Analysis

- This study deals with the detailed performance analysis of various machine learning classification techniques on feat boost feature selection algorithm using 5 widely used datasets.
- We compared the feat-boost algorithm with PSO and genetic algorithm. Featboost gives good results.
- We used both Random Forest and SVM algorithms for software fault prediction and found that Random Forest performed better.
- We searched for the industrial applications of Software Fault Prediction that are currently being used.
- We also found some of the possible industrial applications of SFP.

Industrial Applications



- **Microsoft**: Microsoft has used to predict software faults in its Windows operating system. By analyzing historical data on software faults, the company has been able to identify the most relevant metrics for predicting faults, such as the number of code changes made to a particular module.
- **IBM**: IBM has developed a software fault prediction system called the IBM Fault Analyzer that uses feature selection algorithms to identify the most relevant metrics for predicting faults. This system has been used to analyze large software systems and identify potential faults early in the development process.
- **NASA**: NASA has been using software fault prediction techniques to ensure the reliability of its spacecrafts. In particular, the agency has developed a suite of tools for software fault prediction and diagnosis that uses feature selection algorithms to identify the most relevant metrics for predicting software faults. These tools have been used in numerous space missions, including the Mars Rover missions.

Possible Industrial Application

1. Aerospace and Defense: Fault prediction can be used to ensure the reliability and safety of critical systems, such as those used in aircraft and defense equipment.

2. Healthcare: Fault prediction can be applied to medical software systems to prevent errors that could put patient safety at risk.

3. Banking and Finance: Fault prediction can help identify potential security breaches and prevent financial losses due to software failures.

4. Manufacturing: Fault prediction can be used to monitor and maintain complex machinery and prevent costly downtime due to equipment failures.

5. Energy: Fault prediction can be applied to energy production and distribution systems to ensure the safety and efficiency of the infrastructure.

6. Transportation: Fault prediction can help prevent accidents and improve the safety of transportation systems such as trains, buses, and autonomous vehicles.

Conclusion

- In conclusion, software fault prediction using feature selection algorithms is a valuable approach for improving software quality and reducing maintenance costs.
- Feature selection provides benefits such as improved accuracy, reduced dimensionality, enhanced interpretability, and effective resource allocation.
- Further research and advancements in feature selection have the potential to make a significant impact on industrial software development and maintenance processes.

References

- Smith, J., Johnson, M., & Brown, A. (2018). Fault prediction in open-source software using FeatBoost algorithm. Journal of Software Engineering, 10(2), 87-104.
- Johnson, M., Smith, J., & Anderson, L. (2020). Predicting software defects using FeatBoost algorithm in an industrial context. Proceedings of the International Conference on Software Engineering, 135-142.
- Das, S., & Konar, A. (2008). Software fault prediction using AdaBoost algorithm with decision trees as base classifiers. Journal of Systems and Software, 81(5), 649-660. doi: 10.1016/j.jss.2007.05.022.
- Zhang, X., Han, W., Liu, J., Liu, X., & Cai, Y. (2017). Software fault prediction based on FeatBoost algorithm. Journal of Ambient Intelligence and Humanized Computing, 8(4), 567-576.

THANK YOU!