Defect Prediction: Using Machine Learning

Kirti Hegde, Consultant
Trupti Songadwala, Senior Consultant

Deloitte
Abstract

Defect prediction using machine learning serves to improve the software quality by formulating predictive classification models from different code attributes to enable timely recognition of fault prone modules in the software system.

Machine Learning is the emerging technology which uses computer algorithms to leverage human experience and automate manual efforts to enhance efficiency and cost effectiveness. Machine learning can overshadow the traditional approach of software testing giving rise to defect prediction models to accentuate the flaws in the system.

Defect prediction is the smart way of automating the process of software testing using machine learning technology. Defect prediction model will help in categorizing the different types of defects in a software system. This kind of defect profiling will ensure a quick and more systematic fix as compared to the traditional method of defect detection and fixation.

Integrating machine learning with defect prediction is bound to have a potential impact in industrial practice and pilot testing to the next level in a conglomerate world.
# Table of Contents

Abstract..................................................................................................................... 2  
Table of Contents ..................................................................................................... 3  
Introduction .............................................................................................................. 4  
Challenges in the Traditional Method................................................................. 5  
Development of Defect Prediction Model............................................................ 5  
Evaluation of Defect Prediction Model ................................................................. 7  
Key Points of Data Prediction Models................................................................. 8  
Strategic Planning based on the Model ................................................................. 9  
Author’s Point of View ............................................................................................ 10  
Conclusion .............................................................................................................. 11  
References & Appendix ......................................................................................... 12  
Author Biography ................................................................................................... 13
Introduction

Software testing forms one of the critical components of a Software Development Life Cycle. As the complexity of the application increases, so does the expense of testing. Amalgamation of data science with the latest technology in the smartest way can help in Defect Prediction, which will enhance testing cost-effectiveness.

Machine Learning approach is one of the latest boon of technology which can be used to build predictive models using the data repositories, defect logs from the previous releases of the software system. All the data can be used to identify the components of the application which are most likely to be defect prone.

Basically we are developing a predictive pattern for the defects found in the software, based on the statistical metrics and vectors. This is called as defect prediction using the machine learning.
Challenges in the Traditional Method

In the current global market, we have so many huge software systems that are being maintained from a long duration of time. As the requirements of the software systems keeps on growing so does the complexity of the applications. The critical functionalities of the same system has to be observed and tested over and over again, as a part of regression testing.

Testing the same piece of code repeatedly may sometimes decrease the accuracy of the error detection. With growing complexity, the time, expense and resources needed to test the architecture also increases gradually.
Development of Defect Prediction Model

Defect prediction can be used in profiling defects that might appear in software systems before the new releases. Data repositories like source code, user stories, bug reports from the previous releases serves as the input. Post data collection, metrics from software development life cycle like the cohesion between the modules, complexity ration, and severity of the previous defects is used for developing metrics. Using these parameters a defect prediction model can be build using machine language and the statistics of the data.
Evaluation of Defect Prediction Model

After the collection of the input data based on the requirements, different versions of the model can be created, with the help of machine learning technique. The model can be classified based on the different evaluation factors of the models as follows:

### Number of Bugs uncovered

Each model will differ from each other based on the technique used, and so will the number of the bugs from each of the prediction model. Selection of a model depends on the efficiency of the model. For instance here we can select the second model, as this uncovers the maximum number of bugs in the system, as compared to the other models.

### Performance of the Model

The performance of the model can be measured or evaluated based on the number of actual defects against the number of predicted defects for each of the module. Based on the comparison ratio, you can rate the efficiency of the model and select the best suitable model.
Key Points of Data Prediction Models

**Data:** The basic science behind machine learning techniques is the analysis of large amount of data. These data can be further used for Data Replication Studies for various different studies involving different data sets.

**Metrics:** The metrics used to derive the various data models can further enhance in inferring the current market trends, which in turn will help leverage the business comprehensively. Such data sometimes helps in leveraging the quality of software and the practices used in software development and testing in general.

**Models:** Developing data prediction model requires historical data to a large extend. For applications that are built for the first time and newly released applications, when the availability of the data is limited, data can be shared from cross projects of similar kinds. This can give insights to some hidden aspects of the different systems.

**Performance:** The statistical measures used in deriving the model, helps in evaluating the precision and performance of the models. This also helps in deriving the efforts required to address the defects found in the system.
Strategic Planning based on the Model

Based on the detailed study and analysis of the Defect Prediction model, we can draft the graph for the expense/cost of the testing required for the business at each level of complexity of the system. This method is highly effective in the strategic Test Planning.
Author’s Point of View

Defect prediction using machine learning represents the next advanced step in the software testing milestone, this is where digital technology embraces and imitates human comprehension to create value.

The Raising Future Trends

Numerous fortune 500 companies and capital firms have already recognized the significance of this trending technology in business and have started making strategic investments to enable new potentials. The market for predictive analysis using machine learning is growing rapidly.

Impacting the Value in Market

This technology when augmented with human skills can create a huge vortex across the global markets with cost savings and automating efficiencies. As per market analysis and research, over the next 5 years, more than half of the software industry will incorporate predictive analysis built on cognitive computing functionality.

Customization of the Tools

The whole idea of defect prediction works best when it is used to customize existing tools in order to enhance to improvise the client experience engagements. The market is predicted to grow 2.5X faster than the traditional tools for software testing using machine learning.
Conclusion

Defect prediction model will help the identification of risk areas of software systems at an earlier stage in the software development life cycle. This will enable developers help leverage the code even before the testing stage. As the defect prediction model helps avoiding major risk areas in the software system.

The few challenges faced in the Defect Prediction at this point in time is the availability of the historic data and the accuracy of the prediction pattern for the first time of implementation.

The defect prediction model has more advantages over disadvantages. It can be used for forming algorithms for the different activities performed in testing through identifying the most suitable and correct patterns. Various black box techniques can be improvised using the defect predictive models.

The prediction models can be developed in separate and flexible packages, so that a single model with customizations can be used across different software systems with maximum benefits and minimum cost and efforts.
References & Appendix

https://en.wikipedia.org/wiki/Machine_learning
https://en.wikipedia.org/wiki/Software_defect_indicator
https://www.scientific.net/AMM.687-691.2182
https://www.weforum.org/agenda/2017/01/lifelong-machine-learning/
http://www.softwaretestingclass.com/software-estimation-techniques/
Author Biography

Kirti Hegde has over 6 years of Software Testing experience with specialization in functional and mobile testing for Web, Salesforce and Mobile Applications. She is an ISTQB Certified Software Testing Professional with expertise in Microsoft and HP Testing Tools. She is currently on an assignment with a public sector client and is responsible for the functional and automation testing activities for the solution.

Trupti Songadwala is a dynamic professional with around 9 years of experience on Business Intelligence and Data Warehousing in specific and Management of team. Technical Forte in Teradata - Experienced in ETL Development with domain experience on Retail, Banking and Public sector. She is currently working in development and enhancement through learning the Work fusion RPA (Robotic Process Automation).
THANK YOU!