**Cover Letter**

# Team

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Student ID** | **Email** | **Github Username** |
| Iknoor Singh Arora | 40082312 | [iknoorcan123@gmail.com](mailto:iknoorcan123@gmail.com) |  |
| Sukhpreet Singh Bhatia | 40083564 | [sukhpreetbhatia025@gmail.com](mailto:sukhpreetbhatia025@gmail.com) | Ssdev27 |
| Aditi Bhayana | 40083419 | [aditibhayana029@gmail.com](mailto:aditibhayana029@gmail.com) | Aditi2904 |
| Aakash Ahuja | 40082822 | [aakashahuja1993@gmail.com](mailto:aakashahuja1993@gmail.com) |  |
| Ashmeet Singh | 40070369 | [ashu6811singh@gmail.com](mailto:ashu6811singh@gmail.com) |  |

* Replication Package

<https://github.com/aditi2904/MeasurementProject>

Software Measurement Metrics Report

Aditi Bhayana

*Master of Applied Computer Science Concordia University*

*Montreal, Canada aditibhayana29@gamil.com*

Aakash Ahuja

*Master of Software Engineering Concordia University*

*Montreal, Canada* [aakashahuja1993@gmail.com](mailto:aakashahuja1993@gmail.com)

Iknoor Singh Arora

Master of Applied Computer Science, Concordia University

Montreal, Canada [iknoorcan123@gmail.com](mailto:iknoorcan123@gmail.com)

Ashmeet Singh

Master of Software Engineering Concordia University

Montreal, Canada [ashu6811singh@gmail.com](mailto:ashu6811singh@gmail.com)

Sukhpreet Singh Bhatia

Master of Applied Computer Science

Concordia University

Montreal, Canada [sukhpreetbhatia025@gmail.com](mailto:sukhpreetbhatia025@gmail.com)

***Abstract*—Software measurement is a quantified attribute of a characteristic of a software product or the software process .It is one of the most important parts in the Software Industry. The software is measured to anticipate future qualities of the product or process to enhance and maintain the quality of a product or process. In order to determine the importance of the measurement techniques in Software Engineering, we have measured five different metrics on four different open source projects, and provided the correlation analysis on as much metrics as possible for each project.**

***Keywords—software, engineering, measurement, metrics, code coverage, test suit effectiveness, complexity.***

* 1. INTRODUCTION

The industry uses the Software Metrics in order to improve the decision making during the design phase as well as the development phase of the software. Different metrics may have a different role for the measurement of different kind of Software. With only useful metrics of the software, we can achieve the best results in effectiveness and efficiency. Since the same metric may behave differently in different kind and volume of the software, sometimes it may be trivial to determine the perfect metric to measure the software

As the software evolves, the code and the architecture become more complex, which can inevitably induce bugs in the system? Finding bugs later in the developmental stages can increase the cost of software remodeling. Software developers usually use various metrics for determining the quality of the system. Using these metrics for predicting design defects during development as well as every stage can increase the efficiency of the overall development process.

With Software Metric, the managers in the software development team often get the clear idea of return on investment, different aspects of improvement; to plan about the productivity and effort of the resources, schedule management and overall cost. Calculating different software metrics helps the entire team in their own ways. For example, the managers can benefit from the metrics to prioritize and track the issues such a way that they can improve the productivity of the team. It can lead to delivery of quality software product. Just like the managers, developers can also benefit from the measurement metrics. For example, the development team often use the code coverage or product related metrics to communicate the status/quality of the Project. In Software Maintenance, changes made on a system after its delivery takes up to 90% of all total cost of a typical software. Introducing new functionalities, altering bugs, and changing the code to improve its quality are major parts. There has been much research concentrating on the study of bad design practices, also called smells, defects, ant patterns or anomalies in the literature The calculation of the metrics often has multiple interpretation and/or definitions and/or calculation techniques. It might lead to inconsistency and lack of understanding of the metric for its purpose. The metrics can be computed more accurately if they are calculated by the development team. The metrics must have several important characteristics, such as, Simple and computable, Consistent and unambiguous (objective), Use consistent units of measurement, Independent of programming languages, Easy to calibrate and adaptable, Easy and cost-effective to obtain, Able to be validated for accuracy and reliability, and Relevant to the development of high-quality software products

Here is a list of metrics that we will discuss further in this paper:

1. Test Coverage Metrics: Statement and Branch Coverage
2. Test Suit Effectiveness: Mutation Score
3. Complexity Metric: McCabe Complexity by calculating Cyclomatic Complexity
4. Software Maintenance Metric: Adaptive Maintenance Effort Model
5. Software Quality Metric: Post-release Defect Density

In this paper, we will also perform the correlation analysis amongst different kind of metrics. The correlation coefficient is a measure of linear association between two variables (here in our case, it will be the values of metrics). We have calculated the Spearman as well as Pearson correlation to find the correlation amongst these metrics.

* 1. TEST SUBJECTS
* *Apach*e *Commons Net*

**Size:** 63k LOC

## Version Control System: Git

**Issue Tracking System:** Jira

**Project:** <https://github.com/apache/commons-net>

**Bug Tracking:** -<https://issues.apache.org/jira/projects/NET/issues/NET-679?filter=allopenissues>

**Latest-Version:3.6**  
Apache Commons Net library implements the client side of many basic Internet protocols. The purpose of the library is to provide fundamental protocol access, not higher-level abstractions. Apache Commons Net library contains a collection of network utilities and protocol implementations. Supported protocols include: Echo, Finger, FTP, NNTP, NTP, POP3(S), SMTP(S), Telnet, Whois.

* *JFreeChart*

**Size:** 297k LOC

**Version Control System:** Git

**Issue Tracking System:** Github

**Project** - <https://github.com/jfree/jfreechart>

**Bug tracking** - <https://github.com/jfree/jfreechart/issues>

**Latest Version: 1.5**

JFreeChart is a comprehensive free chart library for the Java(tm) platform that can be used on the client-side (JavaFX and Swing) or the server side (with export to multiple formats including SVG, PNG and PDF). The library is licensed under the terms of the GNU Lesser General Public License (LGPL) version 2.1 or later.

Here are the five different metrics that we considered to measure the five test subjects mentioned above:

III. METRICS

*C. Mutation Coverage Metric*

Software testing is an important part of the software engineering lifecycle process. It plays a pivotal role in the quality assurance of the software product. During software maintenance, testing is a crucial activity to ensure the quality of program code as it evolves over time. With the increasing size and complexity of software, adequate software testing has become increasingly important. In order to measure the effectiveness, the quality of test suites developed is measured. Good software testing has software tests than detect and find faults, if a test cases is unable to detect fault it will be hard to build a good software. This is fulfilled by Mutation Testing.

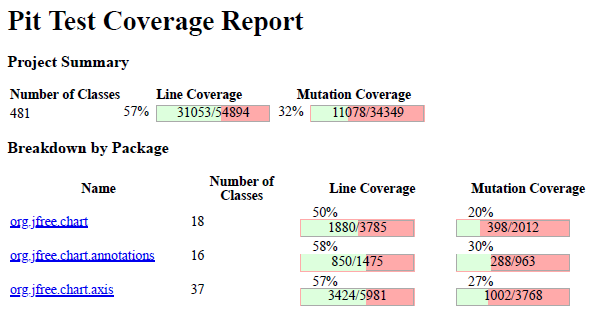
It’s a Whitebox-testing used to develop the hidden test cases by modifying code and check if it could detect the error probability. Furthermore, changes in code statements are made at the lower level so that the meaning of the code remains neutral. The changed statements are the called the mutant and the score derived is known as the mutation score. Thus, to know the failed mutant codes, the statement should be vigorous.

Mutation Score can be **defined** as the number of Mutants that were killed divided by the total number of Mutants multiplied by 100.

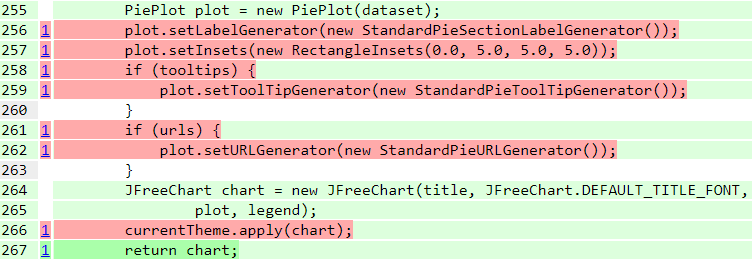
**Collecting Data**

We can collect the data for mutation testing using Pit-Test tool. This tool can be installed in Eclipse by the name of Pitclipse. To execute this tool, we need to run the project under PIT configuration of the PIT test configuration. The results are finally generated in Html file which is called the Pit report.

Sample Pit report:



Mutations performed during Test Coverage:



1. *Cyclomatic Complexity Metric*

In the software development process, there are many difficulties to analyze quality of code. To measure one kind of quality fo this task is Cyclomatic complexity. There are fundamentally two ways to identify the complexity. One among them is the essential complexity which inevitable. The second one is the accidental complexity which is the byproduct of development done in hurry. This can be measured using McCabe complexity metric which is also known as Cyclomatic complexity. This is one of the simplest metric to calculate accidental complexity.

The mathematical method to make the program measured is also done by Cyclomatic Complexity. This was suggested by Thomas McCabe, he said that all the control flow graphs have a fixed number of independent paths. He also proposed a formula to calculate the complexity

**CC = E - N + 2P** or **CC = D + 1**

*Where,*

*E is the number of edges in the graph  
N is the number of nodes in the graph  
P is the number of connected components in the graph  
D is the control predicates in the graph*

Pros of Cyclomatic Complexity:

● It calculates the total possible linearly independent paths executed for a component.

● The number of independent paths can be used to determine the number of test cases needed to cover the component.   
  
Cons of Cyclomatic Complexity:

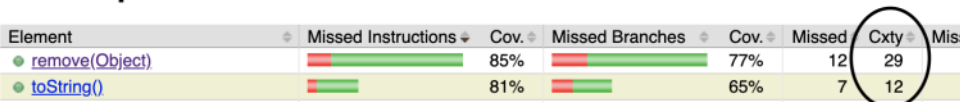
● It considers all the control predicates as the same level of complexity. E.g. for loop and if statement can have different levels of complexity.

● It doesn’t take the nesting into account when calculating from the independent paths.

● It does not consider the flow of data for the component.   
There are several approaches to measure the project and one of the simplest yet a powerful tool to measure. Among the structural or the quantitative approach, quantitative is simple counting of code lines and the project size. This simple approach applied under the different schemes can be very useful to analyze the complexity of the project.

According to the JaCoCo documentation, “JaCoCo also calculates Cyclomatic complexity for each non-abstract method and summarizes complexity for classes, packages and groups”. We used the same configuration as we had for the coverage metrics to calculate the complexity. In the JaCoCo report, we can see the Cyclomatic complexity field along with the other coverage metrics.

Cons: The approach we follow to calculate the complexity has its own drawbacks. Fundamentally this metric considers all the predicate as one, which does not take in account complexities of loop statement as compared to the if statement. This also means that the metric does not take into account the nested loops. The metric calculates multiplicity as a linear component.

In this project, the Cyclomatic complexity metric is calculated using the Jacoco plugin. As described in the JaCoCo documentation, JaCoCo calculates complexity for each non-abstract method and summarizes complexity for classes, packages and groups. Below is the screenshot of the acv generated by the JaCoCo plugin where the covered complexity column represents the Cyclomatic complexity given class wise.

1. *Software Quality Metric (Post Release Defect Density)*

Defect is what is the deviation observed from an expected behavior. Once the software is released to the customer and then if defects are encountered, is referred as post-release defects.

It is defined as  the number of defects confirmed in software/module during a specific period of operation or development divided by the size of the software/module.

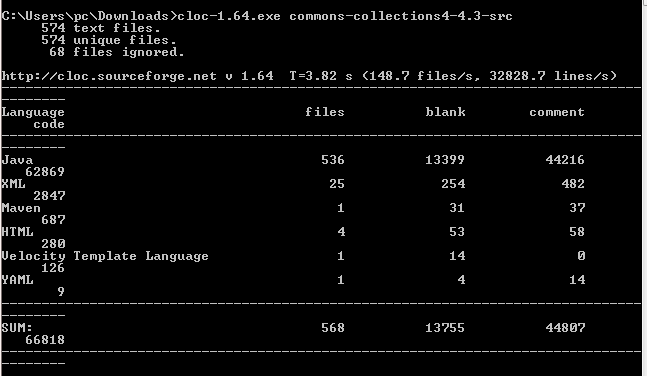
It is counted per thousand lines of code also known as KLOC.

Defect density is used as the indicator for the Product Quality.

## Defect Density = Defect Count / Size of the Software

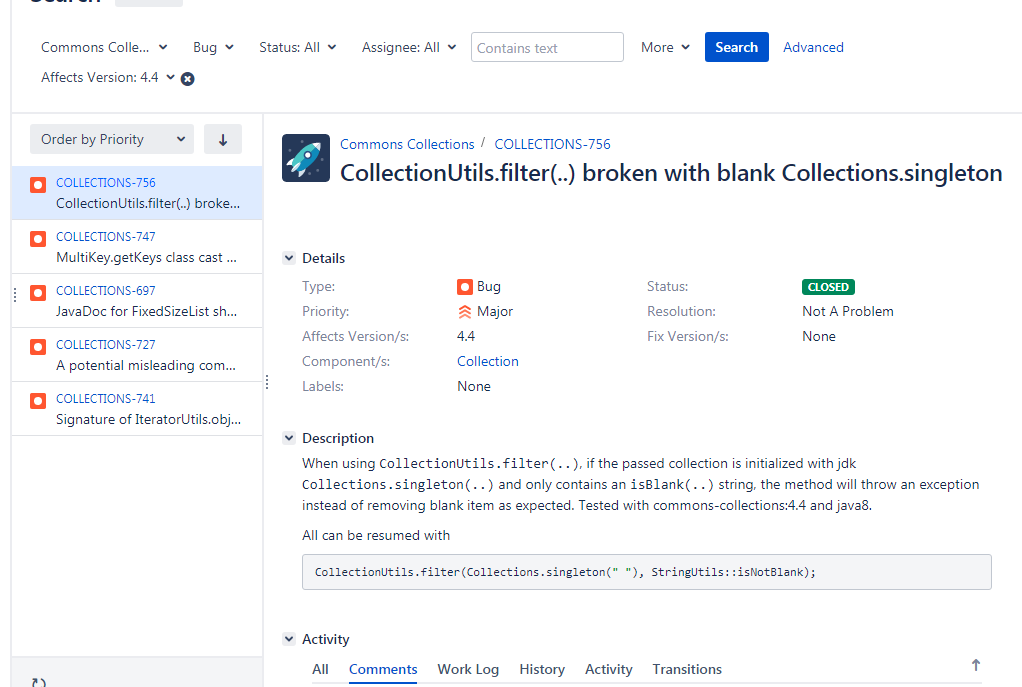
The defect density can be calculated by the type of issues found in the issue tracking system of the Projects. We only take in consideration the issues found out by the in-house team of the project which they track using their respective issue tracking system. This metric also depends on the quality of the testing team as well.There are multiple kinds of issue tracked using JIRA or other tracking systems Like Wish or New Feature, which are not bugs, but the features or pre-release issues that developers usually work on. Other being the tasks, which are part of the Developers development requirement. Each individual developer is assigned a task to work on. Another kind of issue is bugs. They are generally detected when the stories are deployed to the production environment and the testing team is testing for the runtime defects in the system. If any anomaly is found, it is tracked by identifying it as a bug. We also need the source lines of code to calculate the size of the system. There are many tools available to gather this data. But we have used cloc [(https://github.com/AlDanial/cloc/](https://github.com/AlDanial/cloc/)). To run cloc on Windows computers, one must first open up a command (aka DOS) window and invoke **cloc**.exe from the command line there.

The data collection using this tool is illustrated as below:

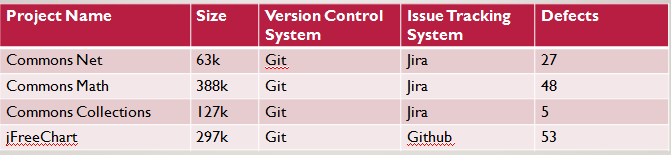


[Figure 6] Calculating the source lines of code

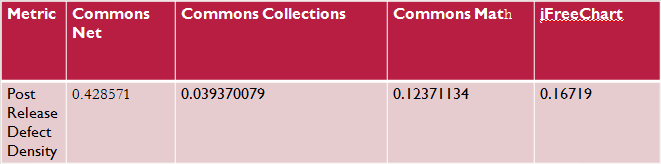
We also need the number of bugs to calculate this metric. We can calculate the number of issues for different versions using their respective Jira Reports from the user interface and then export them as a CSV from the easier calculation. In order to achieve this, we first have to filter the issues on JIRA Tracking System by the type “Bug” (the one with red icon). And we can apply another filter with the version number and then we can see the list. With the option on the top right corner, we can export the list to the CSV for further calculations. Below Figure Illustrates Jira Report after filtering out Bugs for a particular Version of a Project.



[Figure 7] Jira Report

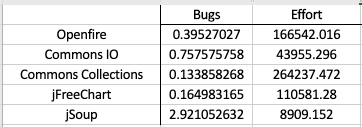


[Table ] Defects for all the Test Subjects



[Table ] Post Release Defect Density for all the Test Subjects

As we can see from the table above, the number of bugs is not always propositional to the size of code. It can be propositional to the size between two versions in the same project. But we cannot generalize that assumption. Hence, we can conclude that the number of bugs in the system does not solely depend on the size of the project. There are other factors as well. One of the factors can be the complexity as well as the quality of the development team.



[Table 6] Results for Post release defect density and Maintenance Effort

From the table, we can see that the more effort the team makes in maintaining the software the less bugs we can encounter after the release. In the table, the column bugs represent the defect density. We can strongly suggest that maintain the software system is an important aspect in order to improve the system with respect to bugs and many other aspects. We achieved the strong Spearman correlation of –0.9 and the Pearson correlation of -0.717431289 during our analysis which shows the importance of the maintenance efforts in the software development.

## Observations on Correlations

 While higher test coverage might have the better test suite effectiveness for most of the projects, but it cannot be generalized to all the systems and teams. It also depends on the quality of the test suits and experience of the developers in automated testing.

 Cyclomatic complexity helps determine the complexity of the system with respect to the coverage of the system. If we are able to test more components in the system, the code becomes less complex.

 Even though the correlation between cyclomatic complexity and coverage metrics is better than the test suite effectiveness and coverage metrics. Both the metrics cannot be under-estimated or over- estimated. They have an important role to play in their own respective ways.

 Quality of the software depends on the quality of the entire team. It does not just depend on the coverage of the tests or the quality of development team.

 In large project, if the testing is not part of the project from the beginning, it is harder to identify the bugs from the development teams and the post release defects may increase exponentially. Even though they start testing the code towards the end of development, it is difficult to cover the entire code before the release and measure the software automatically. Such large projects require more time into production; hence it pushes the deadline and the delivery is not efficient.

 After the software is deployed, it is a crucial to maintain the software in order to improve the quality of the software. The maintenance of the software often leads to less bugs in the production environment.

* 1. TOOLS AND CALCULATIONS

1. *JaCoCo*

The JaCoCo is developed by EclEmma team. It is a popular tool to generate the coverage reports of the Java project. It is available for most of the build tools used in Java (Maven,

Gradle, etc.). It is also available for the IntelliJ IDEA IDE as a plugin. We can enable this plugin and use it by configuring the run configurations of the tasks in IntelliJ. It uses many automated testing frameworks when calculating the code coverage such as jUnit, etc. And it supports the large number of versions of these testing frameworks as well. The configuration steps are mentioned along with the replication package.

1. *CLOC*

We have used cloc tool to calculate the Source Lines of Code. It is developed by AlDanial and hosted on github. We can download this tool in all the operating systems. Since we are using MacOS Mojave, we installed it using brew package manager in darwin environment using the following command:

brew install cloc

We can use this tool using the command line interface of the operating system. All we have to do is go to the root directory of the project and run “cloc .” command and it will calculate the SLOC for us.

This tool supports many programming languages. The detailed guideline to use this tool is available here [(https://github.com/AlDanial/cloc/](https://github.com/AlDanial/cloc/))

1. *PIT-Testing tool*

We have used PIT test tool for calculating mutation coverage in our project. PIT is a state-of-the-art mutation testing system, providing gold standard test coverage for Java and the jvm. It's fast, scalable and integrates with modern test and build tooling. The reports produced by PIT are in an easy to read format combining line coverage and mutation coverage information.

We can configure the PIT test tool as a plugin to the maven or gradle repositories. Once we add the maven dependency of PIT test tool, the IntelliJ IDEA automatically imports the plugin binaries to make it run with the test task of the project.

1. *Correlations (Microsoft Excel)*

We have used Microsoft Excel to perform the calculation on the data that we have gathered using the above tools. Microsoft Excel has built-in functions to calculate the correlation between two data sets. We can use the RANK.AVG function to calculate the rank and CORREL function to calculate the correlations. The detailed configuration steps are mentioned along with the replication package.

* 1. RELATED WORK

 According to paper, “Coverage Is Not Strongly Correlated with Test Suite Effectiveness” who tests on Java projects there is a low to moderate correlation between coverage and effectiveness when the number of test cases in the suite is controlled for. In addition, they found that stronger forms of coverage do not provide greater insight into the effectiveness of the suite. Their results suggest that coverage, while useful for identifying under- tested parts of a program, should not be used as a quality target because it is not a good indicator of test suite effectiveness.

 For most reasonable definitions of coverage metrics, the effort invested in deriving the metrics and measuring them pays off in the form of better error detection. The value of increased confidence in the design’s correctness almost always outweighs the overhead of measuring coverage. Because this confidence depends on the connection between bug classes and coverage metrics.

 According to paper, “Coverage Is Not Strongly Correlated with Test Suite Effectiveness” The coverage of a test suite is often used as a proxy for its ability to detect faults. However, previous studies that investigated the correlation between code coverage and test suite effectiveness have failed to reach a consensus about the nature and strength of the relationship between these test suite characteristics. Moreover, many of the studies were done with small or synthetic programs, making it unclear whether their results generalize to larger programs, and some of the studies did not account for the confounding influence of test suite size.

 The cyclomatic complexity is the commonly used metric in software engineering. It is based on the control flow structure of the program. Each procedure’s statements are considered as a graph. And the complexity is measured by linearly independent paths inside the graph.

 There has been a significant number of effort estimation models in practice. The commonly referred is the Constructive Cost Model, which supports the cost estimation, effort, and timeline for the project. Some model suggests the use of estimated cost or/and the algorithmic cost and/or the function points as we covered in the lecture. The relevant papers related to such work cited in the references.

 With the growing era of Machine Learning, there are recently surfaced models that could be used to measure the efforts in Software Maintenance. Gray and MacDonell used ML-based algorithms such as Artificial Neural Network, Regression Analysis, and a measurement model based on function point, and also the fuzzy logic. With all their analysis, they concluded that the model that uses fuzzy logic proved to perform well with various inputs, but it gave a bit poor accuracy compared to the Artificial Neural Network. And the research has shown that the stepwise regression with bi-directional elimination method provides better accuracy than the linear regression.

 According to the research work, “Estimate the post- release Defect Density based on the Test Level Quality”, the author had a focus on today’s software industry, that concentrates on enhancing the quality of a software product, which can be done during testing and development phase by improving the code quality. But to measure the effectiveness of code, post release defect density is considered an important metric that can be calculated by author “as the number of detected defects per 1000 lines of source code.” Then it was concluded that the post

release defect density of a software is decreased to half of its value, when test level quality is increased from lower level. Thus, they follow an inverse relation and hence, defect density is evaluated as an important factor in determining intrinsic product quality.

* 1. SUMMARY AND CONCLUSION

We have used five test subjects to calculate the correlations amongst the measurement metrics. Based on the observation from the results, we conclude that a single metric or a specific metric is not suitable to give the accurate estimate of the system. These techniques of the measurement provide the overall idea about the system. We also need to take an account of internal product measurement as well as the measurements based on the nature and experience of the team. However, the metrics calculated using the software measurement techniques are also reliable in their own respective fields.

* 1. REFERENCES

1. Laura Inozemtseva and Reid Holmes, School of Computer Science University of Waterloo Waterloo, ON, Canada. Coverage Is Not Strongly Correlated with Test Suite Effectiveness

[http://www.linozemtseva.com/research/2014/icse/covera](http://www.linozemtseva.com/research/2014/icse/coverage/coverage_paper.pdf) [ge/coverage\_paper.pdf](http://www.linozemtseva.com/research/2014/icse/coverage/coverage_paper.pdf)

1. Serdar Tasiran, Compaq Systems Research Center Kurt Keutzer University of California, Berkeley. Coverage Metrics for Functional Validation of Hardware Designs [https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumb er=936247&tag=1](https://ieeexplore.ieee.org/stamp/stamp.jsp?tp&arnumber=936247&tag=1)
2. Thomas J McCabe. A Complexity Measure. I”, IEEE

Transactions on Software Engineering.

1. Ayman Madi, Oussama Kassem Zein and Seifedine Kadry. “On the Improvement of Cyclomatic Complexity Metric”. International Journal of Software Engineering and Its Applications
2. Jane Huffman Hayes, Sandip C Patel, Liming Zhaom, “A Metrics-Based Software Maintenance Effort Model”. Eighth European Conference on Software Maintenance and Reengineering, 2004. CSMR 2004. Proceedings.
3. Vinke, L. (2011). Estimate the post-release Defect Density based on the Test Level Quality. Retrieved February 09, 2019, from [https://research.infosupport.com/wp- content/uploads/2017/08/MasterThesis-LammertVinke- Final.pdf.](https://research.infosupport.com/wp-content/uploads/2017/08/MasterThesis-LammertVinke-Final.pdf)
4. What Are Software Metrics and How Can You Track Them? <https://stackify.com/track-software-metrics/>
5. Importance Of Software Measurement And Metrics. [https://www.ukessays.com/essays/information- technology/importance-of-software-measurement-and- metrics-information-technology-essay.php](https://www.ukessays.com/essays/information-technology/importance-of-software-measurement-and-metrics-information-technology-essay.php)
6. JaCoCo Documentation [(https://www.eclemma.org/jacoco/trunk/index.html](https://www.eclemma.org/jacoco/trunk/index.html))
7. PIT Test <http://pitest.org/>
8. Project Proposal of Team K