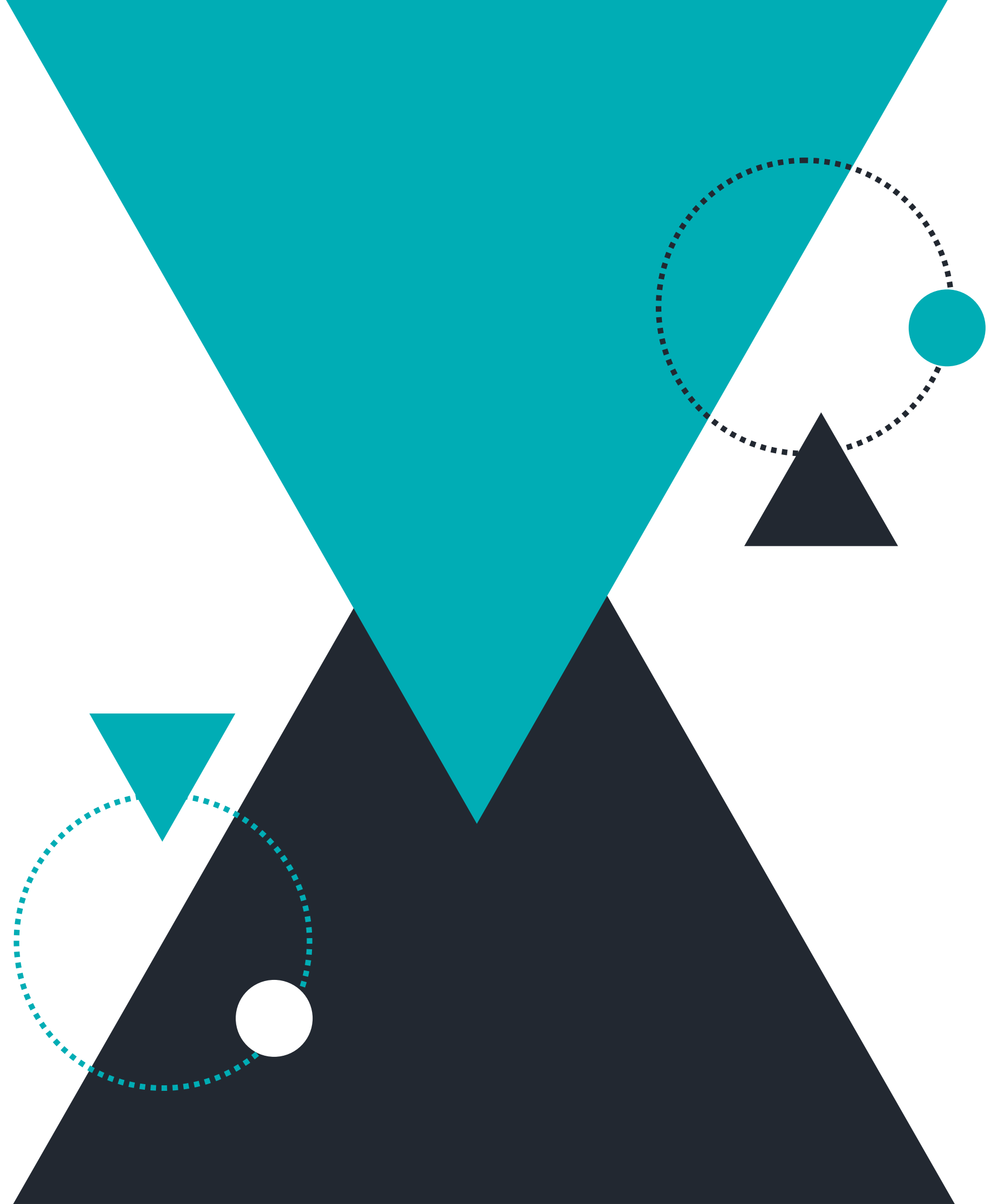




MADE BY : TUSHAR ARORA (20514803119) & HARSHIT SAINI (20214803119)

SOFTWARE FAULT PREDICTION

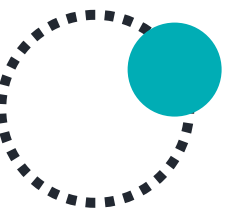
Mentored By : Mr. Varun Goel & Mr. Sachin Garg





Abstract

The IT and software industry has grown tremendously over the past few years, creating an increasing impact on the lives of people and on society as a whole. Consequently, we must make the software and applications more accurate, free of major errors, and more reliable. Therefore, predicting software flaws could be very useful in the IT field and will have a profound impact on society at large.



The topics we have learnt so far

01

Python & Jupyter Notebook

We went over Python fundamentals and became acquainted with the Jupyter notebook environment in order to run code easily.

03

Research Papers

We have also gone through a couple of research papers that have given us great insights about different algorithms and data sets that are great for solving this problem.

02

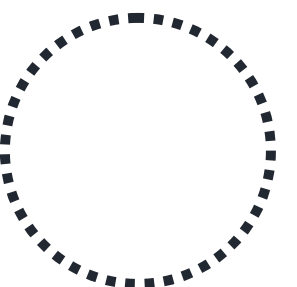
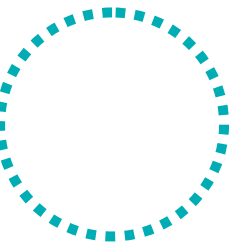
How to use libraries

Next, we have learnt about different Python libraries like Numpy, Pandas, Scikit-learn, Matplotlib, etc., so that we can easily code the algorithms and plot the results.

04

Analysing Datasets

At last, we gathered a bunch of data sets that are open source and publicly available.



What we are doing next

The following are the topics that we are going to learn so that we have a good grasp on the concepts that are used in the project.



Machine Learning

In the next phase, we will be starting with the fundamentals of machine learning and its different concepts.



ML Algorithms

We will also be learning different algorithms like linear regression, KNN, etc.



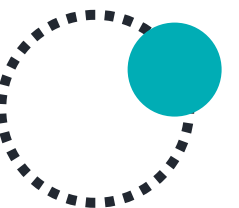
Result Plotting

Next, result plotting is the main thing for finding how the algorithms and models are working.



Performance Analysis

In addition, we will learn techniques for determining the performance of our model and comparing it to existing datasets and models.



About DataSets

In this experiment, we have used 3 open source publicly available data from PROMISE Software Engineering Database. These datasets Tim Menzies et al. have been used in their research paper [1]. In another study, Jureczko et al. [2] have been assembled a software fault prediction model to predict the software defects using machine learning algorithms. They have discussed in their paper about 8 projects (PROMISE Repository) data and by taking 19 CK metrics and McCabe metrics for constructed a predictive model. In our study, we have used 22 attributes for building our automated fault predict model. Table 1 shows 22 different attributes from software defect datasets including 21 independent metrics and one is outcome information. i.e. which is faulty and no-fault.

We are using **JM1**, **CM1**, **PC1** datasets which were implemented in C language.

Reference: [1] T. Menzies, J. Distefano, A. O. S, and R. M. Chapman, "Assessing Predictors of Software Defects." [2] M. Jureczko and L. Madeyski, "Towards identifying software project clusters with regard to defect prediction," in Proceedings of the 6th International Conference on Predictive Models in Software Engineering - PROMISE '10, 2010, p. 1.



About DataSets

Table 2: Details about datasets

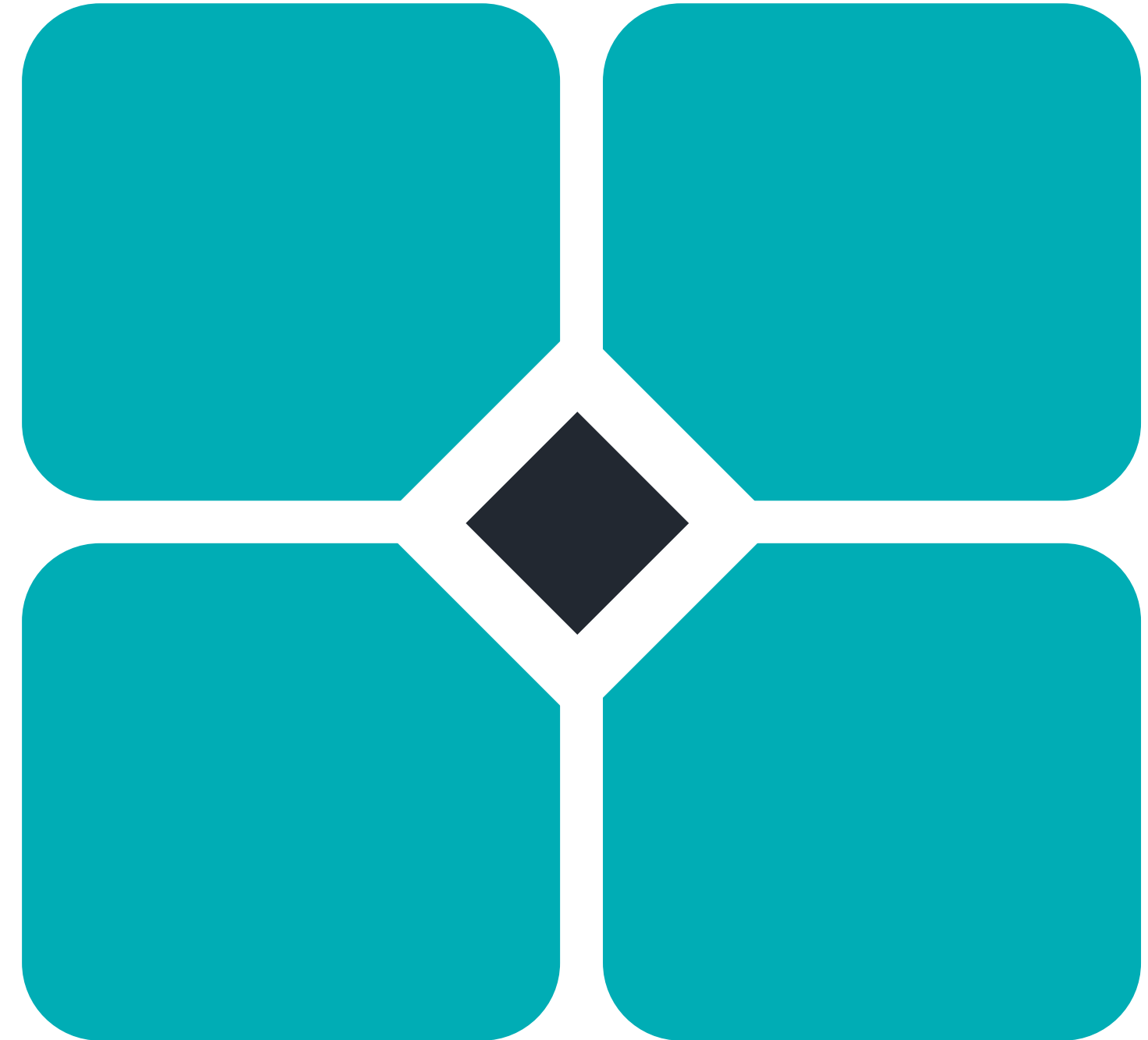
No	Dataset	Missing	Instance	Class distribution	
		attribute		True	False
1	JM1	None	10885	8779 (80.65%)	2106 (19.35%)
2	CM1	None	498	49 (9.83%)	449 (90.16%)
3	PC1	None	1109	1032 (93.05%)	77 (6.94%)



Conclusion

So finally, we are going to conclude with some points that we will be considering once we are done with running algorithms on the datasets mentioned.

- ▶ In the next part, we will be analysing the performance of the model and the techniques that we will be using in this phase.
- ▶ We will also be comparing the results with the other results already obtained with the algorithms.
- ▶ Next, we will be running the same model that we will be making with other datasets to verify if the model is working with accuracy or not.





SOFTWARE FAULT PREDICTION

THANK YOU

Have a great day ahead.

