**AN EXPERIMENTAL STUDY FOR SOFTWARE QUALITY PREDICTION WITH MACHINE LEARNING METHODS**

**Abstract:**

Successful implementation of a software product entirely depends on the

quality of the software developed. However, prediction of the quality of a software

product prior to its implementation in real-world applications presents signiﬁcant

challenges to the software developer during the process of development. A limited

spectrum of research in this area has been reported in the literature as of today.

Most of the researchers have concentrated their research work on software quality

prediction using various machine learning techniques. Another aspect pertaining to

software quality prediction is that the prediction must be achieved in the earlier stages

of software development life cycle in order to reduce the amount of effort required

by the developer in course of the development of a software product. In this paper,

we carry out a comprehensive review of machine learning techniques which have

been used to predict software quality

Software quality estimation is an activity needed at various stages of software development. It may be used for planning the project`s quality assurance practices and for benchmarking. In earlier previous studies, two methods (Multiple Criteria Linear Programming and Multiple Criteria Quadratic Programming) for estimating the quality of software had been used. Also, C5.0, SVM and Neutral network were experimented with for quality estimation. These studies have relatively low accuracies. In this study, we aimed to improve estimation accuracy by using relevant features of a large dataset. We used a feature selection method and correlation matrix for reaching higher accuracies. In addition, we have experimented with recent methods shown to be successful for other prediction tasks. Machine learning algorithms such as Xgboost, Random Forest and Decision Tree are applied to the data to predict the software quality and reveal the relation between the quality and development attributes. The experimental results show that the quality level of software can be well estimated by machine learning algorithms.

Existing System:

Software applications may contain defects, originating from requirements analysis, specification and other activities conducted in the software development. Therefore, software quality estimation is an activity needed at various stages [1]. It may be used for planning the project based quality assurance practices and for benchmarking. In addition, the number of defects per unit is considered one of the most important factors that indicate the quality of the software

There are two directly comparable studies on software quality prediction using defect quantities in ISBGS dataset. In the first study, the two methods (MCLP and MCQP) were experimented with the dataset and the results were compared [3]. The quality level was classified according to: number of minor defect + 2\*number of major defect + 4\*number of extreme defect. The quality of level was to be either high or low. They used k-fold cross-validation technique to measure MCLP and MCQP’s performance on the ISBSG database. Release 10 Dataset (released in January 2007) which contained 4,017 records and 106 attributes was used. After preprocessing, 374 records and 11 attributes remained in the dataset.

Proposed System:

In this project author is using various machine learning algorithms such as Random Forest, Decision Tree, Gradient Boosting, Bagging Classifier, Logistic Regression, BernoulliNB and CNN to predict software quality. To implement this project author has used two datasets but those datasets are not available on internet so whatever dataset student sent I am using that dataset only. This dataset saved inside Dataset folder.

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : MINIMUM i3.
* Hard Disk : 40 GB.
* Ram : 4 GB.

**SOFTWARE REQUIREMENTS:**

* **Operating System:** Windows 8
* **Coding Language**: Python 3.7

CONCLUSION

In this paper we have experimented classification algorithms using Scikit-learn library on two dataset. We have experimented with recent algorithms that support multi-class classification. The accuracies achieved by using these algorithms are 92.28% on EBSPM Dataset and 92.22% on ISBSG Dataset. In comparison to previous directly comparable studies, acceptable level multiclass quality prediction could be achieved.

Successful implementation of a software product entirely depends on the

quality of the software developed. However, prediction of the quality of a software

product prior to its implementation in real-world applications presents signiﬁcant

challenges to the software developer during the process of development. A limited

spectrum of research in this area has been reported in the literature as of today.

Most of the researchers have concentrated their research work on software quality

prediction using various machine learning techniques. Another aspect pertaining to

software quality prediction is that the prediction must be achieved in the earlier stages

of software development life cycle in order to reduce the amount of effort required

by the developer in course of the development of a software product. In this paper,

we carry out a comprehensive review of machine learning techniques which have

been used to predict software quality