ANALYSIS OF SEVERAL CLASSIFICATION TECHNIQUES FOR DETECTING FAULTS IN DIFFERENT KIND OF SOFTWARES

[1]**Ritika , **[2]**Er. Saurabh Sharma**

[1]Research Scholar,[2] Assistant Professor

Department of CSE

SSU, Palampur

ABSTRACT: An important aspect of examination process of software is the flaw spotting. A software blemish is basically an error, bug , mistake in software that cause it produce damage outcome or answer and deteriorates the software quality. A testing process composed of numeral of step to find fracture in software .In detection of mistake different techniques used to detect the existence of faults .In this reappraisal paper, various sorting technique as Neural Network, SVM classification, Random Forest, Naive Bayes, Decision Tree reviewed for null functional requirements. The techniques are reviewed on the footing of certain parameter. In this paper we review various classifiers and their classification techniques to find which classifier is better in detecting the faults.

KEYWORDS- Classification, Null Functions, KNN, SVM, Decision tree, Naive Bayes, Software fault detection, Random forest

I. INTRODUCTION

Software fault detection is an crucial step during testing process of computer software [1]. Defect is an erroneous belief, bug in software that cause it produce wrong outcome or unexpected solvent. As software play key task in the today living so software system should be fault free.A faulty software cannot take better results and will deteriorate in execution due to presence of unnecessary faults. Many Software applications with existence of faults can results many critical job. For a better software there is always a greator need to diminises the faults, computer error and failures of an software. There is always greator need of identifying mistake in under developing software. In software system the fault detection is an crucial step of testing stage of SDLC. It identifies or detect the error prone or defected modules [1].

Software Testing

Examination process analyses demeanour of the product to detect conflict between existing products and required conditions. Testing is process to identify completeness, rightness and effectiveness of an software. It check if predicted or expected result peer with actual result [2]. There are techniques for detective work faults as unit, integration testing and many more [3].

Software Defect

An flaw, failure, error in computer software program known to be defect. An software defect cause it to fruit unexpected outcome or produce wrong result. The error in software may be resulted due to mistakes in programs code or in its model. Software defects leading in wrong outcome which further results in degrading software quality. Testing ensures quality of delivered cartesian product and quality product must be delivered to the customers.

Software Defect Prediction

SDP identifies modules of software that are error prone or defected. Designation of mistake results in price reduction and time reduction of developing software. SDP is an major step of examination phase.SDP includes software metrics, their attribute such as transmission line of the code etc. The intention of lens of blemish foretelling model includes ordination new modules of software accordingly defect-proneness and classifying them whether it is new software or not. The SDP process consists basically two piece:

1. Data Collecting

2. Model Construction

First of all, data is collected from old software modules with known defect numbers. Furthermore, model based approach like machine learning toward construct model and obtained data from them. The proposed simulation predicts defects in new module. An sort of enormity has been received according to forecasting model and allocates resource accordingly order of testing [4].

In ranking the SDP predict which module have more bugs to define software quality enhancement [5]. In ranking task the SDP design purpose is to predict the congeneric defect numbers, in maliciousness that estimation of precise number of defects is better than estimation the rank of faculty, because it gives more information than the ranks [6][7].

Classification

Classification, also known as categorization, is basically an project of automatically applying labels to a information, such as emails, network pages or images. Conveying a specified message or e-mail to "spam" or "not spam" class is example of classification.

Classification is supervised learning approach in which computer programs learns from particular facts comment then uses this learning to classify new notice.

Categorizations is a technique where we categorize data into numeral of classes. The classification aspiration towards spot the stratum or class for the given information.

II. RELATED WORK

Software defect predictions models help in identifying and detecting faults present in different modules of software . SDP role modelling used for calibre improvement [1]. In this paper they proposed that mar may lead in deprivation of software quality which will be reason for software failure .Its vital to make drive for minimizing software defects . Although, these efforts can toll money, time and resources. This paper identifies effective factors which in turn helps to improve software program tone and productivity. It also focuses on implementing the various blemish prediction examples for reduction of defect magnitude. SDP model construction can gain by straight

optimizing simulation public presentation measure for ranking task have been discussed in paper [4] .This work is comprised of learn-to-rank strategy towards the defect prediction . This work prove that straight boosting model performance measure is effective for the learning-to-rank approach. [6] .There is an basic need of building a prediction framework (for defects) in the system. Examination conducted via sovereign testing team [4] .Such prediction model serve as early quality indicator and assist examination team in controlling and also in managing activities of test execution. Here from the preceding phases metrics are obtained are identified and analyzed which is used to consider the capable predictors for constructing such models.

Classifications technique used for sorting aim such as to predict and detect the mistakes in computer software modules. Classification techniques uses different classifiers for this intention. Each classifier having features or advantages also some disadvantages accordingly they are employed. There are certain papers in which different classifiers have compared with another on basis of their parameters. Such as we reviwed paper emphasises on various classification techniques (statistical, machine learning based) and regulate study on each of these techniques[8]. Data mining has wide orbit applications, it integrates proficiency from various field (artificial intelligence, machine eruditeness, spam filtering, Network encroachment detection, pattern recognition) for analysing huge quantity of information. Sorting techniques usually powerful in modelling communications. Each method can used in various situations as needed where one tends to be useful while the other may not and vice-versa. These categorization techniques are implemented on several data sets .Hence they illustrate how data to be determined and grouped for a new set of information. Every technique having some characteristic and restrictions as in this paper. Based on that conditions, corresponding performance and feature one can be selected. Accordingly in another paper [9] they compared and analyse functioning of some classifiers based on accuracy, execution time, type of dataset and domain. The analysis and comparison shows that k-star having highest accuracy for larger dataset but other are not and for smaller dataset performance of algorithms are comparatively same. Therefore no exact algorithm is best suited for specific situation, their performance depend on the type and size of data sets. One algorithm is more appropriate for one data set while other is not appropriate for the same data set.

SVM, KNN algorithmic program are explained and evaluated to conclude which one having good accuracy and in which conditions in paper[10]. SVM is a representation for measurements for performing supervised erudition. SVM technique work for data analysis and patterns recognition. SVM used for categorization labor. Similarly KNN is also sorting algorithmic program, although sorts data using training examples. Here two different classifiers used for image classification. With a proper observation we have concluded KNN in general having good predictive accuracy in small proportion also they use high memory , and are difficult to interpret. Overall after execution found that KNN is a quit good classifier but SVM classifier does not sustenance multiclass classification i.e only supports binary classification. So SVM accuracy is less as compared to KNN in many executions. [11]This paper does comparison of three classifiers(linear SVM, WKNN, decision tree (DT)) based on their functioning by applying facts from optimized and non-optimized sensor set solutions. These algorithmic program skilled with well-known information and later examined with unknown information for detecting fault. The performance of the SVM and WKNN was similar. The paper found that DT classifier performance is reasonably worse for lower berth sample sizes used for preparation phase and SVM is slower in face of training and examination times compared to WKNN and DT.

Classifiers

An algorithm used to implement sorting or organization particularly in a concrete implementation, known as classifier.
There are different classifiers such as:

Support Vector Machine

SVM is an supervised learning.SVM is basically employed for sorting and regression problems. SVM classifier performs classification through forming hyperplanes in multidimensional space. It handles categorical variables and multiple continuous variables. SVM is an representation of data in form of points in space, these points are broken into classes by a wide gap.

Features and Advantages

Work well for linear separable data. Having Good accuracy.Better search time.Well Suited for smaller dataset.

Disadvantages

More speed and size is required. High complexity. Bad performance for the larger datasets.

Decision Tree

DT classifier is simple classification method. It builds classification model in form of tree structure. In decision tree data set is broken into smaller subsets where tree is incrementally build up at same time. The concluding outcome is a tree with decision, leaf node (represents a sorting or decision)[12].

Features and Advantages

DT algorithm is simple and fast.
It takes less memory.

Disadvantages:

Long training time

Naive Bayes

NB classifier depends on Bayes theorem. Its works on the principle that every span of features being classified is independence of other. Naive bayes classifier employment well in many real-world situation(document classification and spam filtering).

Features and Advantages

Simple to implement. Good computational efficiency and classification rate.Predicts accurate results .

Disadvantages

If data is less then preciseness of algorithmic program decreases. It require large records.

Random Forest

This classifier is assembly learning scheme for sorting purpose. It is set of multiple decision trees, an meta-estimator fits number of decision tree and control over-accomodation.

Features and Advantages

It works efficiently on larger datasets. Providing an estimation of significant variables in classification. It handles thousands input variables and without their deletion. **Disadvantages**
Overfit to those datasets with noisy sorting/ regression tasks.

K-Nearest Neighbour

It is an simple classifier. It is an non-parametric scheme applicable for the categorization task. KNN is instance based learning (lazy learning).KNN classifier uses the neighbour points to predict target class. K-Nearest Neighbour sorting an supervised classification: it stores all objects and find similarity to classify new object. An instance is sorted by neighbours majority vote(the "k" is number of neighbour it checks).

Features and Advantages

KNN is having high accuracy .Easy to implement. Robust to the noisy information. It takes less computational time throughout training phase. Suitable for the multi modal classes. Fast training.

Disadvantage

In larger dataset excessive time is required to find nearest neighbour **.**Sensitive to noisy data.

Neural Network

It is an robust classifier for supervised learning. Its classification is fast but training is slow. It is an network of neurons, arranged in the levels, tarns forming an input into several output. This network is trained by back propagation algorithm i.e output of previous level supply the input for the next level but no feedback to the previous layer.

Features and Advantages

Easy to adjust with few parameter**.** No need of reprogramming. Applicable to real life problems.

Disadvantages

Entails elevated processing time for large network. Difficult to find necessary neurons layers. Yield more time during Learning

III. CONCLUSION & SCOPE

This paper emphasises on different classification techniques used for detecting faults of software. It is concluded that classification techniques are that which classify given data into certain classes. In this review, certain classification techniques such as SVM , ANN, Random forest ,Naive bayes , KNN are reviewed for classification purpose. Each classifier having different features or advantages and also some disadvantages. Based on conditions and features any classifier can be selected and worked with, and KNN having more compensation above other in detection of faults. After reviewing various techniques we concluded that KNN classification technique gives more accurate result for detection of faults as it deals with robust data and multimodel classes.. In future KNN classification technique can be designed for better detection of faults having superior accuracy.

REFERENCES:

[1] Mrinal Singh Rawat, Sanjay Kumar Dubey , "Software Defect Prediction Models for Quality Improvement: A Literature Study", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 2, September 2012 ISSN (Online): 1694-0814

[2] John E. Bentley, Wachovia Bank, Charlotte NC, "Software Testing Fundamentals—Concepts, Roles, and Terminology", 2009

[3] Gaurav, Kestina Rai "Software Testing Techniques for Test Case Generation" International journal of Advanced Research in Computer Science and Software Engineering 2013 pp 261-265.

[4] Xiaoxing Yang, Ke Tang, Senior Member, IEEE, and Xin Yao, "A Learning-to-Rank Approach to Software Defect Prediction", IEEE TRANSACTIONS ON RELIABILITY, VOL. 64, NO. 1, MARCH 2015

[5] Shaik Nafeez Umar, "Software Testing Defect Prediction Model- A Practical Approach", IJRET: International Journal of Research in Engineering and Technology, Volume: 02 Issue: 05 | May-2013

[6] Muhammad Dhiauddin, Mohamed Suffian, Suhaimi Ibrahim, "A Prediction Model for System Testing Defects using Regression Analysis", International Journal of Soft Computing And Software Engineering (JSCSE) e-ISSN: 2251-7545 Vol.2,o.7, 2012

[7]N.Kalaivani1 , Dr.R.Beena2," Overview of Software Defect Prediction using Machine Learning Algorithms" in International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 3863-3873.

[8]Nikam S. S. A Comparative Study of Classification Techniques in Data Mining Algorithms. Orient.J. Comp. Sci. and Technol;8(1)

[9] Ritu Sharma , Mr. Shiv Kumar , Mr. Rohit Maheshwari," Comparative Analysis of Classification Techniques in Data Mining Using Different Datasets " in IJCSMC, Vol. 4, Issue. 12, December 2015, pg.125 – 134

[10]Anchal Tomar , Anshika Nagpal ," Comparing Accuracy of K-Nearest-Neighbor and Support-Vector-Machines for Age Estimation" in International Journal of Engineering Trends and Technology (IJETT) – Volume 38 Number 6- August 2016.

[11] Marcel Jung , Octavian Niculita , Zakwan Skaf " Comparison of Different Classification Algorithms for Fault Detection and Fault Isolation in Complex Systems"

[12]http://medium.com/sifium/machine-learning-types-of-classification-9497bd4f2e14.