Applications of AI/ML in Improving a Company's Efficiency

A DISSERTATION
SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF DEGREE
OF

MASTER OF TECHNOLOGY

IN

SOFTWARE ENGINEERING

Submitted by

Prabhakar

2K20/SWE/16

Under the supervision of

Dr. Abhilasha Sharma

(Assistant Professor)

DEPARTMENT OF SOFTWARE ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

MAY 2022

CANDIDATE'S DECLARATION

I, PRABHAKAR, Roll No. 2K2O/SWE/16 student of M. Tech (Software Engineering), hereby declare that the project dissertation titled "Applications of AI/ML in Improving a Company's Efficiency" which is submitted by me to the Department of Software Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology in Software Engineering, is original and not copied from any source .

Prabhakar

Prabhakar

Place: Delhi
Date: 30/05/2022

CERTIFICATE

I hereby certify that the Project Dissertation titled **"Applications of AI/ML in Improving a Company's Efficiency"** which is submitted by Prabhakar, 2K2O/SWE/16 Department of Software Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology in Software Engineering, is a record of the project work carriedout by him under my supervision. To the best of my knowledge, this work has not been submitted in part or full for any degree or diploma to this university.

Abhilasha Sharma
30/05/2022

Place: Delhi

Dr. Abhilasha Sharma
SUPERVISIOR
(Assistant Professor)

ACKNOWLEDGMENT

The success of this project requires the assistance and input of numerous people and the organization. I am grateful to everyone who helped in shaping the result of the project.

I express my sincere thanks to **Dr. Abhilasha Sharma**, my project guide, for providingme with the opportunity to undertake this project under her guidance. Her constant support and encouragement have made me realize that it is the process of learningwhich weighs more than the result. I am highly indebted to the panel facultiesduring all the progress evaluations for her guidance, constant supervision and formotivating me to complete my work. She helped me throughout with innovative ideas,provided information necessary and pushed me to complete the work.

I also thank all my fellow students and my family for their continued support.

Prabhakar

ABSTRACT

Machine learning is used in several types of problems and performing very well in achieving solutions for those. Similarly, we are going to explore ways in which a software development company can increase their efficiency and enhance product quality by use of this.

A company's efficiency depends majorly upon the workforce and workflow they are using. Maintaining workforce quality and using an efficient workflow can help them in producing quality output.

Voluntary Employee turnover is a great threat for all major companies across the globe as company's overall performance is highly dependent on employee. A lot of investment is done by companies to firstly find a suitable employee and their training according to needs and after all this in retaining those trained and skilled employees in their companies. ML is used to solve the issue of customer churn prediction and to resolve that issue, this study aims to use the same method to predict voluntary employee attrition.

And Software testing is one of the crucial steps in software development and as well as time and energy consuming phase. There are many automation tools like selenium which offers ease to testing phase in many ways but to a limited extent. Also, there are SFP or software fault prediction systems which uses machine learning models and helps in predicting faults' (strength, weakness, opportunity, and threats) analysis is one the first studies we do before starting our research in any field as it clears many myths and doubt before starting and gives a clear view of topic.

CONTENTS

LIST OF FIGURE(S)

LIST OF TABLE(S)

LIST OF ABBREVIATION(S)

NLP: NATURAL LANGUAGE PROCESSING

ML: MACHINE LEARNING

AI: ARTIFICIAL INTELIGENCE

FAQ: FREQUENTLY ASKED QUESTIONS

SDLC: SOFTWARE DEVELOPMENT LIFE CYCLE

KDD: KNOWLEDGE DATABASE DISCOVERY

ST: SOFTWARE TESTING

SFP: SOFTWARE FAULT PREDICTION

1. HR: HUMAN RESOURCE

CHAPTER 1

INTRODUCTION

1.1 MOTIVATION

The ease with which an organization may employ the resources it has available to generate the most goods and services is the efficiency factor. This element can have an impact on both large and small businesses. Large organizations have more resources; inefficiency may not harm short-term results, but it may cause problems overall. Small firms, on the other hand, must be always efficient to survive and expand. The effectiveness of a firm is determined by its employees' commitment to the company's goals and priorities. Staff duties and responsibilities must be clearly defined, and the organization must conduct programs to improve employee abilities. Employees must also be encouraged and motivated by the organization. Business owners must recognize and compensate employees for their contributions to ensure employee satisfaction and productivity. They can also outsource specific operations duties to increase efficiency.

Today a lot of investment is made by organization to drive there experienced and trained employees' number of policies and schemes are made by HR department to reduce attrition is one of the factors that contribute to this. Whenever experience employee leaves a company, it cost both in training and hiring a new employee. Furthermore, when an employee leaves, both tacit and explicit knowledge is lost, and important social relationships may be broken. HR professionals frequently struggle to explain how they add value to their organizations, and one of their responsibilities is to make HR more significant through better decisions. HR departments are increasingly attempting to base decisions on data.

Data-driven decisions can lead to improved organizational performance, which means that if HR departments can make data-driven decisions, they will add value to the organization.

Data analytics refers to the process of making decisions based on data analysis. Businesses use data analytics to interpret and extract information from data that can be used for decision making. Information analytics is changing into additional standard within the field of human resources, and it is the potential to extend the importance of time unit departments among organizations.

Machine learning (ML) may be a standard technique for information analytics prediction. The study of a way to build computers learn from expertise is thought as machine learning. The concept is AN algorithmic rule to find out from information sets and improve as new data is introduced might doubtless be utilized in time unit departments to predict worker

attrition. The problem with this is often that selections created among time unit departments can have extensive consequences for workers, their families, and the organization. As a result, this paper can investigate the chance of victimization machine learning to predict worker attrition. Info derived from information which will be accustomed build selections information analytics is changing into additional standard within the field of human resources, and it is the potential to extend the importance of time unit departments among organizations.

Another way is using ML in software testing process of a company.

- Testing software is costly. Did you know that it can account for up to 25% of overall project costs?

Unit tests and automated integration tests have helped reduce the amount of manual work involved in running tests, but they still require a lot of manual work [4]. You need to create and maintain tests, interpret failed tests, and modify your code accordingly. In addition, current automated testing strategies cannot detect many of the human user interface problems, such as: These include trimmed elements, misalignment, and responsiveness issues. UI testing is a major challenge for traditional automated testing that runs at the code level rather than fully rendered output.

To address this, we propose running and validating tests through the eyes of the user, rather than the code. Visual changes can be detected using computer vision. But what are the rules for filtering out all the insignificant changes, such as when the window size is changed responsively? We believe that machine learning can help to solve some of the above software testing challenges (ML). Because machine learning models can be trained using examples, they do not require a predefined set of rules. Furthermore, machine learning could help to reduce the amount of work and expertise required to write a test, even with modern tools.

1.2 OBJECTIVE

Employee attrition is a challenge that many businesses encounter, with important and experienced staff leaving daily. Many firms across the world are working to eliminate this severe problem. The major goal of this study is to create a model that can help forecast whether an employee would leave the organization. The main concept is to assess the efficiency of employee appraisals and satisfaction levels inside the firm, which can help to prevent staff churn. A new machine learning-based approach was applied to improve various retention strategies for targeted employees. This report also tries to throw some insight on several factors The focus of AI-based testing is to form the testing method a lot of intelligent and economical.

Logic reasoning and problem-solving approaches will be used to enhance the total testing method. Moreover, AI testing tools area unit utilized to execute tests that use knowledge and algorithms to develop and perform tests while not the necessity for human interaction during this testing approach. SWOT analysis is used to investigate a company's internal and external environments by characteristic and analyzing the organization's strengths and weaknesses, more because of the opportunities and dangers to it it's exposed. a vicinity of the target of a SWOT analysis is to assertively verify elements that influence the organization's operation, providing terribly helpful knowledge for strategic planning designing.

1.3 SCOPE

Need of Employee Attrition prediction: -

[1] Managing the workforce: If supervisors or HR learn that some employees are planning to leave the company, they can contact those employees to persuade them to stay or they can manage the workforce by hiring a replacement for those employees [2].

[2] Streamlined pipeline: If all the employees in the present project are working constantly on a project, the pipeline of that project will be smooth; but, if one of the project's efficient assets (employee) suddenly leaves that company, the workflow will be less than smooth

[3] Hiring: If the HR of a specific project learns about an employee who wishes to quit the organization, he or she can regulate the number of hirings and ensure that the important asset is available whenever needed for the efficient flow of work.

Some of the advantages of using artificial intelligence in software testing: -

- Visual validation
- Improved accuracy
- Better test coverage [43]
- Saves time, money, and efforts
- Faster time-to-market
- Reduces defects

1.4 THESIS OUTLINE

There are six chapters in this thesis.

The first is the introduction, which is meant to provide an overview of the work, the problem and the purpose of this study.

The second chapter discusses and analyses the current approach as well as potential alternatives.

The scientific research that underpins the chosen approach will be explained in the third chapter.

The fourth section provides an overview of the AI implementation and the various optimizations performed in the project.

The fifth chapter discusses the state of the art in AI and analyze the work done.

The sixth and last chapter summarizes the acquired knowledge and work that has been done.

CHAPTER 2

LITERATURE WORK

There is numerous research in the topic of employee attrition, and different studies have varying degrees of accuracy and feature sets, making it impossible to determine which study is superior.

There are a few key points from previous research: -

- According to D. G. Gardner and E. Moncarz, giving remuneration appears to be a crucial factor in forecasting employee attrition and performance [29,32].
- S. Kaur discovered that salary is not the only factor affecting attrition in the retail industry; other factors such as workload, performance pay, and a lack of a career plan have all contributed to higher turnover [25].
- Age, tenure, compensation, general job happiness, and employees' development of trust, according to J. L. Cotton, appear to be the strongest indicators of voluntary turnover.

Employee attrition is the leaving company of an employee due to any reason. There can be different reason for an employee turnover like death, retirement, or resignation. These turnovers are categorized into two major types Voluntary and Involuntary. Involuntary turnover referred to those attrition in which an employee is Retired, fired by that company or can be death while voluntary attrition is that in which an employee leaves the company for their own reasons and that reason can be anything like policies of company or job satisfaction. These all leads to reduction in company's workforce. A company cannot do anything for involuntary turnovers (death, retirement) but by a better retention policy it can reduce or even stop the voluntary turnovers(resignations) in their company. Generally, because of voluntary attrition company loses a lot of its high performing employees and that is why they are so much concern about it [26].

To tackle, first we should be aware of the problems and causes that majorly responsible for attritions Attrition can also be reduced if a company is able to predict the vulnerable employee and focusing on those employees more to change their attrition decision .By analyzing the dataset of employees ,their work and their other related details the drivers causing the attrition can be figured out [20] .In this study we are focusing majorly on using machine learning to predict the vulnerable employee and help the HR of company to understand the

major policies or elements needed to be included in their policies so, that they can be retained. This will save their money they might spending in useless retention policies and retaining efficient workforce for their company.

Problem Studied	Data Mining Techniques used	Recommended model	Dataset used
Employee Attrition Prediction [11]	KNN, Naive Bayes, MLP Classifier, Logistic Regression	KNN (Accuracy: 94.32%)	Kaggle sample dataset
Employee churn prediction using SVM following the feasibility study of 4 other algorithms [12]	Naïve Bayes, Support Vector Machines, Logistic Regression, Decision Trees and Random Forests	SVM (Accuracy: 80%)	Particular Sample dataset from HR department of three companies in India
Demonstration of XGBoost against six historically used supervised classifiers and demonstrate its significantly higher accuracy [13]	Naïve Bayes, Support Vector Machines, Logistic Regression, Linear discriminant analysis, Random Forests, KNN, XGBoost	XGBoost (Accuracy: 88%)	Dataset of a certain level of employees in a particular leadership team of a global retailer
Study of turnover prediction models: Logit and Probit models [14]	Logistic regression model (logit), probability regression model (probit)	Logistic regression model (logit) (Accuracy: 90.5%)	Custom dataset drawn from a motor marketing company in Taiwan
Comparison of various decision trees for the analysis of the turnover of employees [15]	ID3 Decision tree, CART Decision tree	CART Decision Tree (Accuracy: 90%)	Kaggle sample dataset
Behavioral comparison of Random Forest and Naive Bayes [16]	Naive Bayes and Random Forest	Naive Bayes (Accuracy: Up to 85%)	Sample dataset of sales agents

Table 2.1 Previous Studies in Employee Attrition [13]

Software is a single word used for many types of applications either web application or mobile apps or desktop software. This software market is expected to grow exponentially in coming years for reference according to statics reports 171 billion mobile apps alone expected to be downloaded in 2022. Also, in a report given by National Association of Software and Services Companies (NASSC), the software industry in India was worth Rs. 243.5 billion or US$ 5.7 billion in 1999-2000 and that is 20 years ago.

In recent year the AI and ML grew exponentially with their applications and usage in industry similarly in software testing there have been many research papers, articles published related which gives various point of views [18].

Some websites and companies also there which focuses on researching the use of AI in software testing and provides various solutions in form of software as a service. Test.ai, functionize.com are some of the most popular websites used which are providing software development companies a feature of executing test cases with machine learning.

Let us understand some widely used AI-based testing techniques in industry.

AI based test cases execution: - In this concept test cases are executed, and results are saved in a database for further learning of machines. So that these learning can be used to predict test cases by itself in future. The main goal of this technique is to reduce or even eliminate the irrelevant test cases from testing which can save a lot of time wasted on these irrelevant test cases.

Another technique used is SFP (software fault prediction) [16].

Software fault prediction: It is known by different names like SFP and software defect prediction, but all are doing same thing which is predicting faulty modulus in the software [20]. This work using machine learning on historical reports of projects and predicting results based on the learning from the data saved in all these years [21].

A lot of research and studies are done on SFP and seen that from 1991 to 2009 the most used classifiers are decision tree and Bayesian learning [18]. The dataset trends of those studies are also shown in graph below [16].

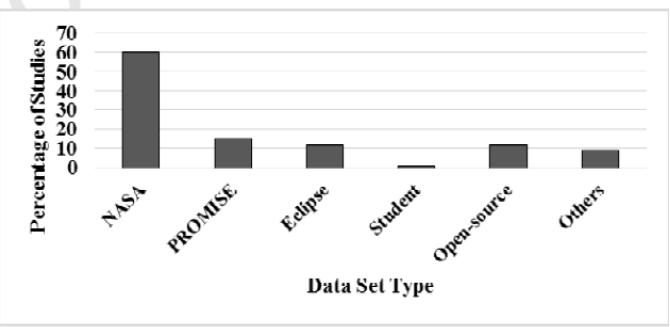


Table 2.2 Trends of Datasets Used [16]

CHAPTER 3

BACKGROUND

3.1 THEORETICAL KNOWLEDGE

In this section a brief outline of some topics that are relevant. These following questions are addressed within the framework of the experiment:

- How can be machine learning can be used in predicting employee turnover?
- What can be different point on using Artificial intelligence in software testing?
- Which metrics are better for prediction of employee turnover?

The first portion discusses the plan of information finding in databases, attended by a definition of dossier excavating and a complete elaboration on machine intelligence usually and relevant algorithms. Following that, the current state of the art in MT and feature study is determined. The portion concludes by defining mechanics proof and lightness the different statuses concerning this.

There are diverse types of SDLC. Some of most used are explained below: -

1 Agile:

Popularized in 2001, Agile gives the ability to create and respond to solutions leading to improvement through collaborative effort of self-organizing and cross-

functional teams and this also help team in delivering a high value product to their customers with fewer headaches in the development process [6]. Its main methodology is managing the project by breaking it into to several phases.

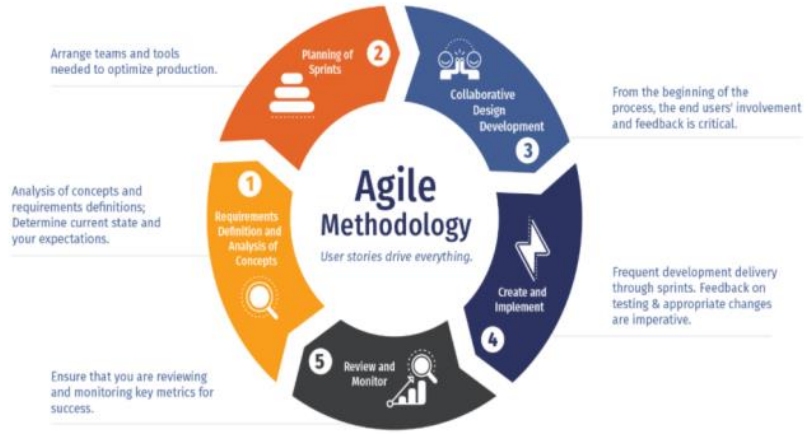


Figure 3.1 Agile methodology [6]

2 Lean Methodology:

This methodology of software development life cycle works on the principle of eliminating or reducing wastes [6][7]. This focuses on delivering product as fast as possible. In other words, this methodology follows skipping less important meetings and less documentation, which proves to be cost effective as well as takes the entire team in decision making process [7]. However, this reduced meeting causes communication barriers sometimes between stakeholders

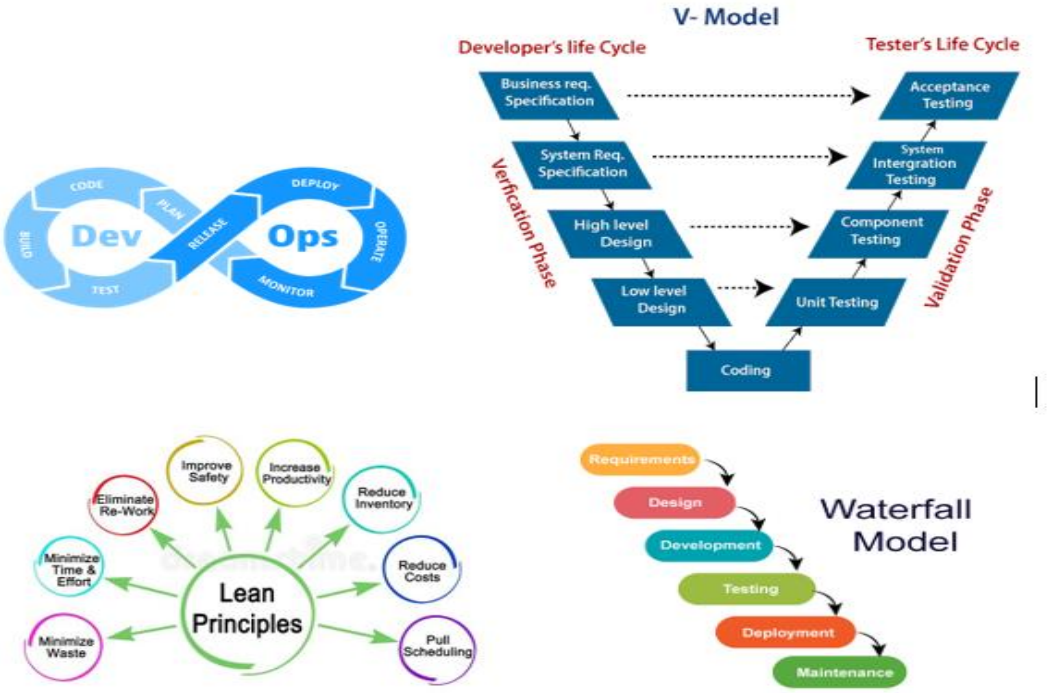


Figure 3.2 SDLC methodologies [6,8,9]

3 Waterfall Methodology:

It is one of the oldest methodologies [8] in the industry and still surviving in industry. This follows a very simple approach of taking one step at a time. Its project development works on principle of completing one phase at a time and when that one phase completes its task then only it is forwarded to next stage with all necessary information from the previous stage [9]. Because of it is this fixed process and rigid nature some experts do not think it as a working SDLC but for the same reason it is best for extremely predictive projects [8].

4 Devops:

It is mainly used by big companies [6][9]. It is considered as hybrid form of both lean and agile methodologies. It is completely a collaborative approach of development team with operating team which give it its name as Devops. Their collaborative work leads to accelerate the entire

process [10]. This methodology is a time saver from an irrelevant communication barrier between two teams.

5 V model:

It is a similar model like waterfall but in V model the testing is done at every stage. Task is given to next stage only after successful testing from the previous stage [10].
Above we have discussed some major SDLC. In all these software life cycles there is a phase called testing phase in which all the products compiled are assessed. Testing is an important phase of SDLC [12].

In testing phase various tests are done on various stages to check diverse types of errors and bugs that can affect the software's efficiency, functionality and can reduces software's reliability.

Software testing have two main categories of testing functional testing and non-functional testing

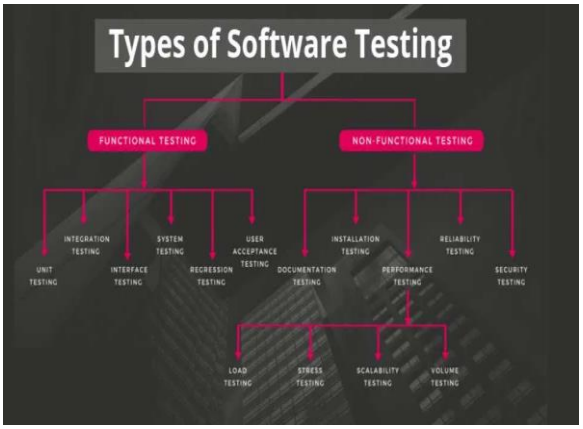


Figure 3.3 Types of Software testing [13]

Functional testing All the functional aspects of a software are tested in functional testing. Functionality of each functional element of software are tested by using them or by running some predefined test cases on them. There is different testing checked in functional testing

- Unit testing
- Integration testing
- End to end testing
- Smoke testing [13]
- Sanity testing [13]

- Regression testing
- Acceptance testing
- White box testing
- Black box testing
- Interface testing

These testing can be done by both automated software and by manual tools. Those tools are decided based on number of test cases and functionality to be tested [14]. Some of automatic tools widely used in industry are JUnit, Watir.

Non-functional testing All the non-functional aspects like usability, readability, security of application is tested in non-functional testing. It is only done after successful functional testing. In Non-functional testing quality upgradation is not limited to user's experience but it used to give software a long-term quality certificate. The non-functional testing is done usually on tools but reviewed manually based on software requirement. In non-functional testing same result can lead to different conclusions for example in load testing high level load for one product may not satisfy the needs in other product [36].

Nonfunctional testing covers several types of testing

- Performance testing [37]
- Security testing
- Load testing
- Failover testing
- Compatibility testing
- Scalability testing
- Usability testing
- Stress testing
- Efficiency testing
- Reusability testing
- Endurance testing
- Disaster recovery testing

Now doing all this testing is not possible and can be a lot time-consuming manually for that various automated tool are used to complete them efficiently. These tools are automated but do not use any considerable machine learning or AI in working.

Some of the most popular tool is Selenium, which is an open source and oldest software testing tool, originally developed in 2004[15]. It is a web application testing tool which give its user a

wide range of testing features along with the recording feature by which after doing the test case tester review it later [16][18]. It is often used for regression testing.

3.2 KNOWLEDGE DISCOVERY

The knowledge discovery of database process (KDD) outlines the total stage of obtaining meaningful information. Although several descriptions of the KDD process, most of them agree on the key components. KDD should be defined as a participatory and iterative process. They lay out nine major steps.

1. Determine the process's aim and obtain any prior knowledge required about the application domain.
2. Select an acceptable data collection from which to extract knowledge.
3. Pre-process the information. This includes deleting noise or potentially hazardous data records, as well as deciding on specific parameters,
   such as how missing attribute values in the data collection are handled.
4. Reduce the features for as by deleting factors that are not relevant to the task.
5. Select a data mining strategy to achieve the KDD process's declared goal.
6. The next work is to select a data mining algorithm after deciding on a broad data mining method. It is vital to remember that this decision is frequently influenced by the end user's preferences, such as whether a comprehensible format or the highest level of prediction quality is preferred.
7. This is the most important data mining stage. The algorithm is then applied to the preprocessed data collection. The algorithm then explores the data for useful information.
8. Interpret the patterns discovered by the algorithm and, if necessary, return to one of the previous phases to change the KDD process configuration.
9. The final phase in database knowledge discovery is to use the interpreted results for other purposes, such as conducting additional study or applying a system to a real-world problem.

   The KDD procedure, as described in step 8, might include numerous iterations and loops. For example, after evaluating the outputs of an algorithm, one can determine that the chosen method was a wrong decision and return to step 5, or that the preprocessing was done incorrectly and return to step 3 after reducing the data to a representable format in step 4.

3.3 DATA MINING

The Data excavating is once more and once more secondhand correspondently with KDD. File excavating may be a return to be disquieted the KDD manner that consists of selecting and requesting the appropriate methodology and set of rules to the straightforward report file [1].

As a result, it is an especially important aspect of the table records locating manner. data processing is that the manner of communicable some variety of file and requesting examine algorithms thereto so as that realize designs or fashions within the straightforward report file, and previous classifying the file into various classifications the employment of these buildings (labels). info structures, enumerations, and sample acknowledgment are with the managed disciplines hid.

Data mining tasks are classified according to the algorithm's understanding of the data set's existing classes:

- Each assignment at which point the development has approach to suggestion and sum values is consider coordinated instruction. These standards are the specific names of the lesson quality, because as proposal standards are the outward truths that the calculation is conceded to utilize, within the way that trait standards and meta file. This demonstrates that the file shape is already celebrated, and the point of these program is to designate unused file to the suitable classes.
- Unlike supervised learning, not the least bit like learning, unsupervised learning includes all errands that do not have get to yield values and during this method endeavor to get structures details by creating categories on their claim [3].

Data mining will be divided into two essential objectives: confirmation and revelation [4]. Whereas confirmation endeavors to demonstrate the user's theory, revelation appearance for already obscure styles at intervals the knowledge. The revelation step is isolated into two parts: portrayal, during which the framework appearance for styles to point out the knowledge in an cheap organize, and expectation, during which the system tries to foresee semipermanent results of knowledge supported styles. The subgroup expectation assignment will be assist divided into classification and relapse tasks. Whereas classification assignments produce settled names, and every data record has one in all these names as its course attribute esteem, relapse errands deliver continuous values.

3.4 MACHINE LEARNING

In the data mining step, it is basic to decide on the acceptable approach for handling the trip fittingly. Usually this can be often consummated through the utilize of machine learning ways. For a protracted time, a serious distinction between individuals and computers has been that folks tend to consequently progress their approach to an issue. Humans learn from their botches and conceive to fathom them by rectifying them or finding higher approaches to fathom the problem. Typical laptop programs do not think about the results of their assignments and

thus cannot create strides their behavior. the sphere of machine learning addresses this explicit issue by making laptop programs that may learn and thus create strides their execution by gathering knowledge [3]. A. prophet was the first research worker to set up a self-learning program in 1952, once he develops whereas classification errands deliver settled names and every info record has one in every of these names as its course quality esteem, relapse errands deliver persistent values.

Machine learning has been employed in data mining, adaptive software systems, and text and language learning disciplines since the 1990s. For example, a computer software that collects data about an e-commerce store's clients and uses this information to make better targeted adverts can learn new things and is close to being artificial intelligence.

Furthermore, machine learning systems are typically classed according to their underlying learning techniques, which are determined by the amount of inference that the computer program can accomplish.

- All classic computer programs employ a process known as rote learning. They do not make any inferences, and all their knowledge must be applied directly by the program because the application is
- All computer programs that can convert information from a given input language to an internal language are classified as learning from instruction. Although the program still has the knowledge of how to do this transformation properly, the computer program must make certain inferences. As a result, this distinguishes a higher degree of learning from rote learning.
- Learning by Analogy, in contrast to Learning through Instruction, aims to acquire new skills that are almost identical to existing skills and hence easy to learn by transforming preexisting data. The ability to create mutations and combinations of a dynamic knowledge set is required for this system. It adds new features that the original computer program did not have, necessitating a great deal of inference.
- Learning from Examples, Learning from Illustrations is right now one among the first common learning strategies because of it gives the first adaptability and grants pc programs to create antecedently obscure abilities or find antecedently obscure structures and designs in data. Learning from illustrations can be a classification and information preparing strategy that predicts the category name of most recent data passages upheld an energetic set of far-famed illustrations. The arranged examination inquiries are tended to amid this work misuse strategies and calculations from this course.

Based on the prediction, in this study we have accessed various supervised learning algorithms to predict the turnover. we are going to use six different classification algorithms, but our focus

will be gradient boosting and random forest as we find these as better from previous work. This section gives detail about the steps followed with dataset and various classification algorithms. The following are brief descriptions of the most common machine learning systems:

3.4.1 Decision Tree

The following area unit transient descriptions of the foremost common machine learning systems: one among the foremost common learning ways could be a call Tree, which could be a classification technique that focuses on Assistant degree simply graspable illustration kind. call Trees create use of knowledge sets created from attribute vectors, that successively contain a collection of classification attributes describing the vector and a category attribute classifying the information entry. a choice Tree is made by iteratively rending the information attack the attribute that separates the information likewise as attainable into the varied existing categories till an explicit stop criterion is met. because of call Trees is simply unreal in an exceedingly tree structured format that humans will perceive, the illustration kind permits users to urge a fast summary of the information.

Nodes in call Trees area unit classified as root nodes, inner nodes, and finish nodes, additionally called leaf. the basis node, which has no incoming edges, represents the start of the choice support method. The inner nodes have one incoming edge and a minimum of two outgoing edges. They embrace a look at supported a knowledge set attribute.

For example, a look at of this kind may raise, "Is the client older than thirty-five for the attribute age?" Leaf nodes contain the answer to the choice drawback, which is usually depicted by a category prediction. a choice drawback may be the question of whether a client in an internet search can create an acquisition, with the category predictions being affirmative or no.

Following a node n, any nodes separated by specifically one edge from an area unit referred to as kids of n, and n is considered the parent of all its kid nodes. a choice Tree is seen in Figure. as an example, a knowledge record with the characteristics chilly, polarBear would be sent right down to the left subtree since his temperature attribute is cold, then to the leaf "North Pole," which might be classed with the suitable label.

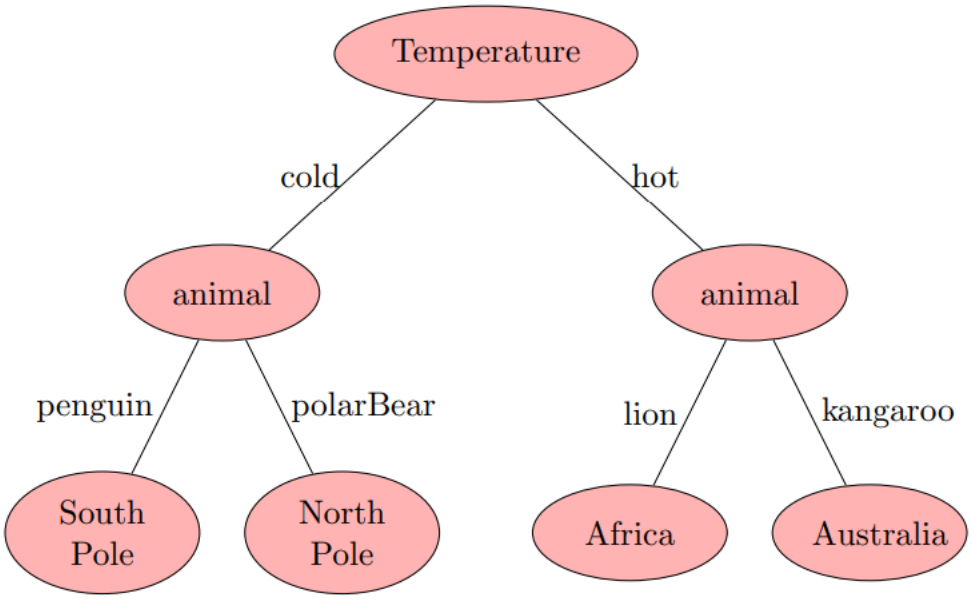


Figure 3.4 Supervised Learning Example

A common information handling procedure is to coach a choice tree, that is by and large utilized for categorization. Its reason is to foresee the worth of a target quality from a bunch of input values. in an administered setting, coaching a choice Tree is completed by finding designs inside the information and building the choice Tree utilizing a coaching set. Taking after that, a gather of antecedently seen tests may be utilized to anticipate the worth of their target attribute. the data records inside the coaching set unit inside the sort of: -

$$\left(\vec{x}, Y\right) = \left(x_1, x_2, x_3, ..., x_n, Y\right)$$

with Y being the target attribute value and ~x being a vector containing n input values, where n is the number of attributes in the data set

A preparation set containing a mark attribute, recommendation attributes, a split test, and a stop test is necessary to train a Decision Tree and therefore produce a classifier. The split test calculates a worth for all traits at the bud. This number displays by virtue of what much news is win by divorcing the bud utilizing this attribute. After that, best choice profit from all attributes is preferred, and the bud is detached into the differing attributes' results. At this stage, the process of deciding the optimum attribute split is recurrent for all established substitute seedlings as far as a stop necessity is join.

The following are stop tests:

• The tree's maximum climax has happened attained.

• The node holds middling records than the allowable minimum.

• In agreements of acquire facts, the best split test does not surpass the threshold.

Mechanized strategy for arrangement a Choice Tree can result in colossal Choice Trees going with parts of diminished categorization control. Trees are as well subordinate on something being overfitted, that riches they are as well around multiplied to the arrangement models. When these trees are utilized to strange file, this leads to frail results. As a result, a trimming strategy has happened conceived. Its point is to expel divisions of the Choice Tree that are less or non-fruitful, within the way that parts based on raucous or off base file or parts that are overfitted. This commonly leads to indeed more prominent veracity picks up and a decrease in bush sum. Since each palpable-globe fundamental archive record holds incorrect or unruly file, this step is exceptionally fault-finding.

Computation Time

The beginning shrub-increasing recipe, that only studies imaginary attributes, appearance a momentary condition of O (m * n2), unspecified area m is that the magnitude of the instructing information set, and n is that the assortment of attributes. The forecast of the well-informed dossier for each attribute takes highest in rank period inside the forest-increasing methods. The principles of the befriended attribute for all information records inside the current instructing set square measure wanted to reckon the dossier gain. The merger of all subsets by any means levels of the choice Tree has an equivalent capacity cause the original instructing to emerge worst-case means. As a result, calculating the dossier gains for each level of the shrub is then O (m * n) in character. on account of calamity-case assortment of shrub levels is n, the issue of instructing a choice Tree is O (m *n2).

After instructing a choice Tree, after step search out use it to foresee classification labels for information records that have nevertheless visible.

To do so, the record is two-passed along from the bedrock bud to a leaf, accompanying each bud's corresponding attribute being proven and again the edges being attended to the appropriate leaf.

Algorithm *Decision Tree Training Process*

1. training set = S;
2. attribute set: A;
3. target Attribute = C;
4. split criterion = sC;
5. stop criterion = stop;
6. *Grow*$(S, A, C, sC, stop)$
7. **if** $stop(S) =$ **false**
8. **then**
9. **for** all $a_i \in A$
10. **do** find a_i with the best $sc(S)$;
11. label current Node with a;
12. **for** all values $v_i \in a$
13. **do** label outgoing edge with v_i
14. $Ssub = S$ where $a = v_i$;
15. create subNode = *Grow*$(Ssub, A, C, sC, stop)$;
16. **else** currentNode = leaf;
17. label currentNode with c_i where c_i is most common value of $C \in S$;

It represents the order of preparing a Choice Tree in fake rule outside communicable into consideration mathematical features. The judgment starts by determining whether the halt need has existed join. If regularly the case, the current hub is named accompanying the first in rank universal consider of all the preparing set's communication names. In case the stop necessity is not join, the computation calculates the part esteem for all traits. The hub is before part into many knots, one each consider of the preferred characteristic. For each fitting subset, the forethought calls the alike make iteratively, holding all information records accompanying the relating consider of the necessary trait.

3.4.2 Artificial Neural Networks

Artificial Neural Networks square measure outlined by Singh and Chauhan as "a mathematical model supported biological neural networks and therefore a simulation of a biological neural system." compared to ancient algorithms, neural networks might handle problems that square measure additional difficult at a lower level of methodology quality [28]. As a result, the key reason to utilize artificial neural networks is their basic structure and self-organizing nature, which permits them to resolve a good vary of problems while not the necessity for more programming intervention. For instance, in a web store, a neural network may be trained on client behavior information to predict whether the user can build an

acquisition.

An Artificial Neural Network is created from nodes, unremarkably referred to as neurons, weighted connections between these neurons which will be adjusted throughout the network's learning method, Assistant in Nursing an activation operate that determines every node's output price supported its input values. completely different layers conjure every neural network. The input layer takes information from external sources, like attribute values from the relevant information entry, the output layer generates network output, and hidden layers connect the input and output layers [20]. The total of all incoming nodes increased by the load of the interconnection between the nodes determines the input price {of every of every} node in each layer.

- Feedforward Networks: - All structures that forbiddance get interpretation from the systematize itself are top-secret as are. This shows that facts stream in one course from the recommendation centers to the yield hubs, pass through ton secret centers. There is no dossier given back to permit the foundation expected fix [24].
- All networks accompanying a response alternative and therefore the skill to talk over again dossier from later stages for the education process in former stages are refer to as Recurrent networks.

Each node's output value is determined by applying all input values to a preset function that is the same for all nodes in the network. The sigmoid function (oj), which is defined as follows, is the most widely utilized function.

$$o_j = \frac{1}{1 + e^{-i_j}} \qquad \text{……(3.3.2.1)}$$

where ij is the result of j's input nodes.

The two primary benefits concerning this work, concurring to Erb, are appeal categorize as being illustration to values betwixt and one and appeal nonlinear character, that frame arrange learning smooth and thwarts burden and dominance assets.

An amazingness impact happens when an unmistakable or any properties have a critical influence the thought point property, rendering distinctive characteristics insignificant and with controlling them [23, p.167]. Figure 5 shows a Feedforward Neural Organize going with three suggestion hubs, a sole concealed coating, and two yield ties within the proposal layer.

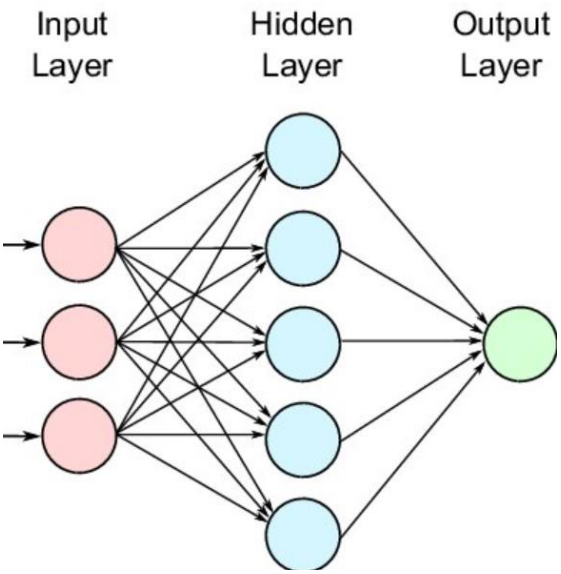


Figure 3.5 Artificial neural Network [24]

ANN instructed utilizing the backpropagation form in a supervised knowledge sketch, which readjusts the weights of the interconnections in the interconnected system established local mistake rates.

Backpropagation

Backpropagation in neural internetworks is that the method of reading just the weights of the interconnections backwards through the neural net mistreatment the network's native error. this suggests that when creating a prediction for a collection of input values, the output worth is compared to the prediction worth, and miscalculation is calculated. This error is then used to reweight the connections, beginning at the sides that are directly connected to the network's output nodes and dealing our manner deeper into it. It is crucial to grasp the first parameters which will be wont to maximize the {training the educational} method once training a neural network

- The education rate shows by means of what speedily the learning assemble is completed activity. The consider of the limit a number 'tween and 1 that's repeated for one nearby mistake each yield esteem. As a result, a knowledge rate of happens in no adaptation. The ideal education rate scene is basic to the knowledge process' ability. If the consider is set also unreasonable, the weights will waver, making verdict ultimate superior principles more troublesome.
- In case the esteem is as well moo, in any case, recognized botches will not have sufficient weight to constrain the arrange into a modern optimization, and the weights will get to be stuck in nearby maxima. A rot parameter can be presented to decide the proper settings.

- Momentum is some other key metric for neural networks. It smooths out the optimization method via way of means of including a fragment of the preceding weight extrude to the contemporary weight extrude.
- The minimum mistakes are a getting to know technique forestall criterion corresponding to the Decision Trees forestall criterion mentioned in part. The getting to know technique is ended every time the network's cumulative mistake falls beneath this threshold.

CHAPTER 4

PROPOSED WORK

4.1 PROBLEM STATEMENT

Software testing is a crucial step in software development but also takes a lot of time and effort. Many automation tools, such as Selenium, make the testing phase easier in many ways, but only to a certain extent. There are also SFP systems, or software fault prediction systems, which employ machine learning algorithms to forecast failures. So, in this article, we will do a full SWOT analysis of the current AI/ML approaches in use in the sector, as well as their benefits and drawbacks. Before beginning, numerous myths and doubts are dispelled, and the topic is clearly defined. To better understand the role of machine learning in software testing, we have included an SFP study in this paper. Voluntary employee turnover is a significant danger to all big firms across the world, as staff effectiveness is so important to the company's overall performance. Companies pay a lot of money to identify a qualified employee, train them according to their needs, and then retain those trained and skilled individuals in their organizations. This study intends to apply the same method to anticipate voluntary employee attrition as it is used to solve the problem of customer churn prediction and resolution. In this research, we use machine learning classification methods with a more complete dataset to improve on previous results in this sector utilizing appropriate feature and model sections.

4.2 PROPOSED METHOD

Employee attrition occurs when an employee leaves an organization for whatever reason. Employee turnover can occur for a variety of reasons, including death, retirement, or resignation. These turnovers can be classified into two categories. There are two types of voluntary and involuntary labor. Involuntary attrition refers to when an employee retires, gets dismissed by the company, or dies, whereas voluntary attrition refers to when an employee leaves the organization for their own reasons, which can include anything from company regulations to job satisfaction. All of this results in a reduction in the company's workforce. Involuntary turnovers (death, retirement) are unavoidable, but a better retention policy can help a company reduce or even eliminate voluntary turnovers (resignations).

To begin, we must first recognize the issues and causes that are mostly responsible for attrition.

Attrition can also be decreased if a company can predict the most vulnerable employees and focus more on them to influence their attrition decision. The drivers of attrition can be identified by analyzing the dataset of employees, their work, and other related details. This research focuses on utilizing machine learning to detect vulnerable employees and to assist firm HR in understanding the important policies or parts that must be included in their policies for them to be retained. This will save them money that they would otherwise spend on ineffective retention practices and keeping productive employee

Strengths Weaknesses Opportunities (SWOT)

SWOT stands for "organized manner of evaluating." A strategy for assessing a project's strengths, weaknesses, opportunities, and risks that exist today and may occur in the future. It can also be used as a foundation for assessing a project's overall potential and constraints. It also serves as a project feasibility study. In general, we aim to factor in all potential hazards and restrictions

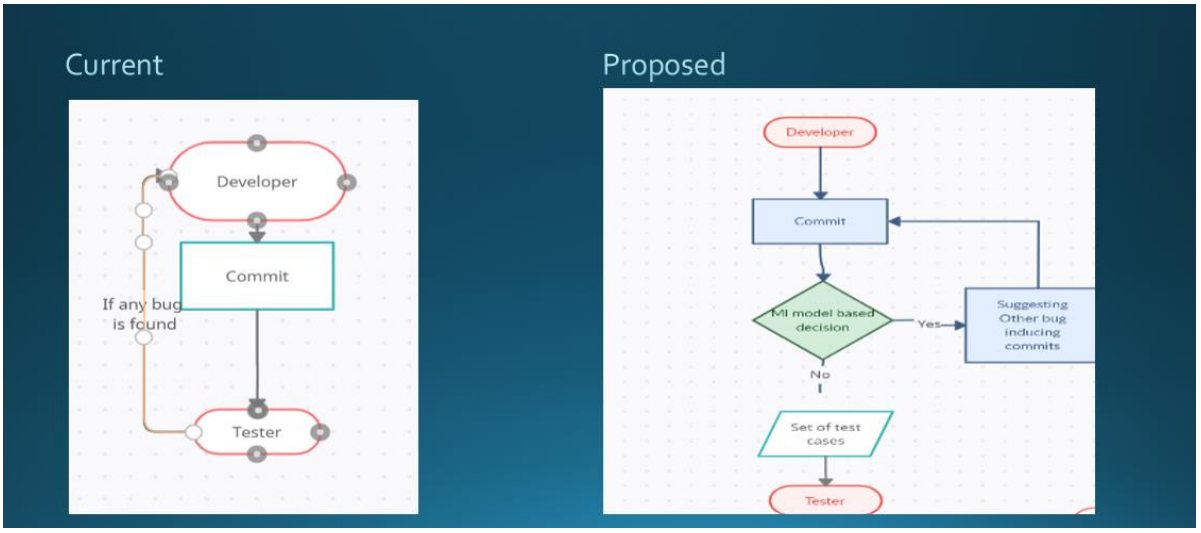


Figure 4.1 AI in Software Testing Flow

CHAPTER 5
WORKING AND ANALYSIS

Let us first understands our dataset used for employee attrition. The dataset used consist of five files. Those files are merged and used as one single file. These files have diverse types of data related to an employee like its designation, gender, age, and all factors by which we will evaluate our result.

Datasets consist of following files:

1. General data
2. Manager survey
3. Employee survey
4. In/out time
5. Data dictionary

I. General data: - This file contains twenty-four unique features and tells general information about an employee. [Age, Attrition, Business Travel, Department, Distance from Home, Education, Education Field, Employee Count, Employee ID, Gender, Job Level, Job Role, Marital Status, Monthly Income, Percent Salary Hike, Standard Hours, Stock Option Level, Total Working Years, Training Times Last Year, Years at Company, Years Since Last Promotion, Years with Cur Manager]. Above are the features in general data file about an employee.

II. Manager survey: - Information like what is an employee's work rating and his/her involvement also matters a lot that is why manager data also matters a lot

III. Employee survey: - This is going to contain an employee thinking and rating he gives himself and his/her work in company like satisfaction from work in his/her life.

IV. In/out time: - We are going to use this to calculate the working hour an employee spends in office.

V. Data dictionary: - when we are using multiple files, we need a common index file which is data dictionary here which tells us about the feature in other files.

All the files of datasets are in CSV format but still a lot of processing is required to make them more suitable for model. Now below the steps taken and followed on dataset.

I. Data processing: -

Data pre-processing or data cleaning is the first step of any machine learning after gathering data. In this step data is made suitable for model removing missing values, null values and making data types suitable by encoding are some common steps done in

different files in dataset.

Also, these all files are merged as one is also a part of data processing step.

II. Feature engineering: -

Data after processing still contain unwanted and less key features which increases the computation time and decreases the accuracy of a model. These irrelevant features impact negatively on the accuracy of the model [29]. In feature engineering select most relevant features we do this by having an overall picture of our data. we use diverse types of charts and bar graphs to analyze our data. we use a correlation matrix two evaluate how features affect each other and selects our best related features only.

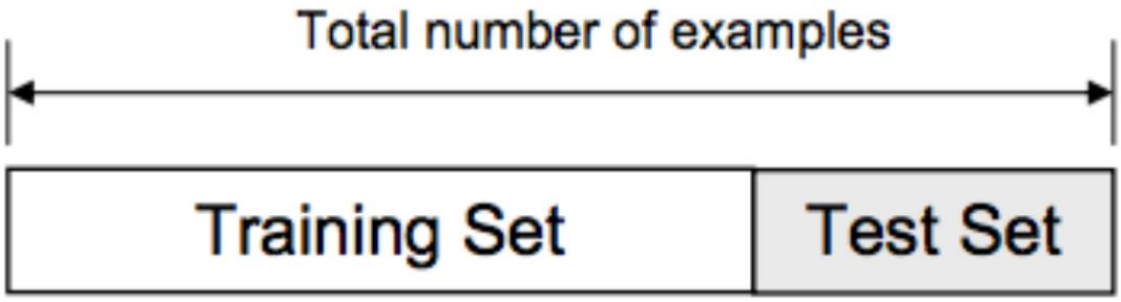
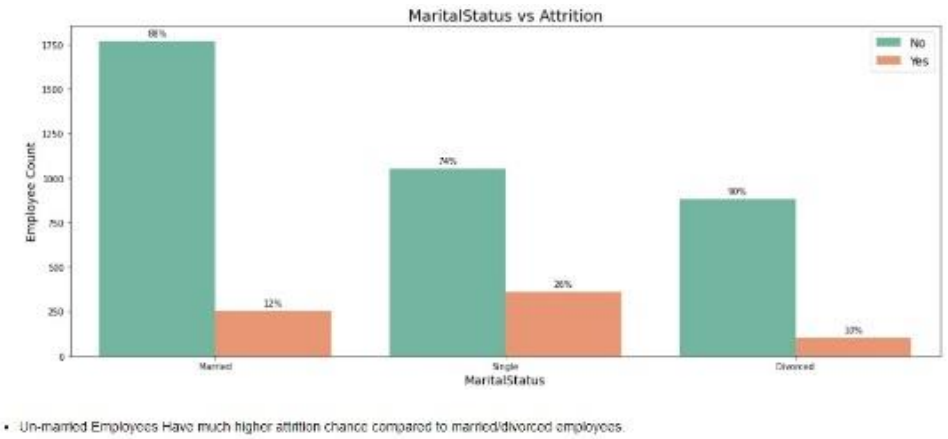
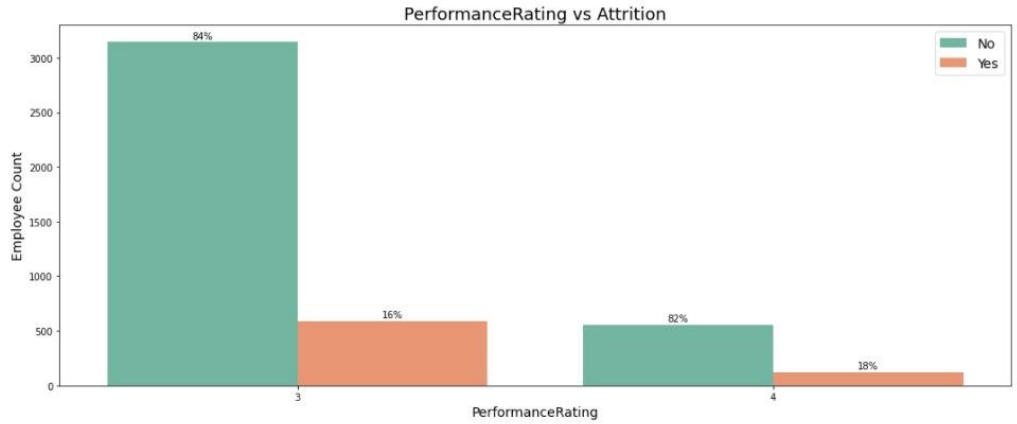


Figure 5.1 Splitting Dataset [27]

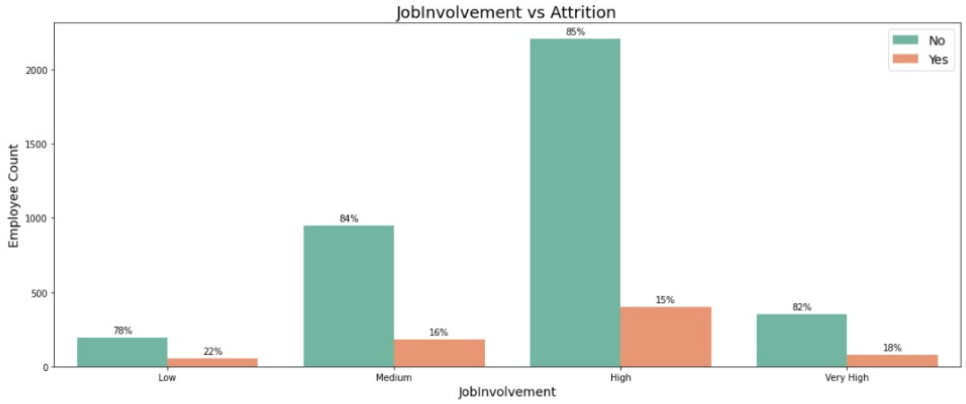
Below shown the results after evaluation of datasets through different exploratory data analysis



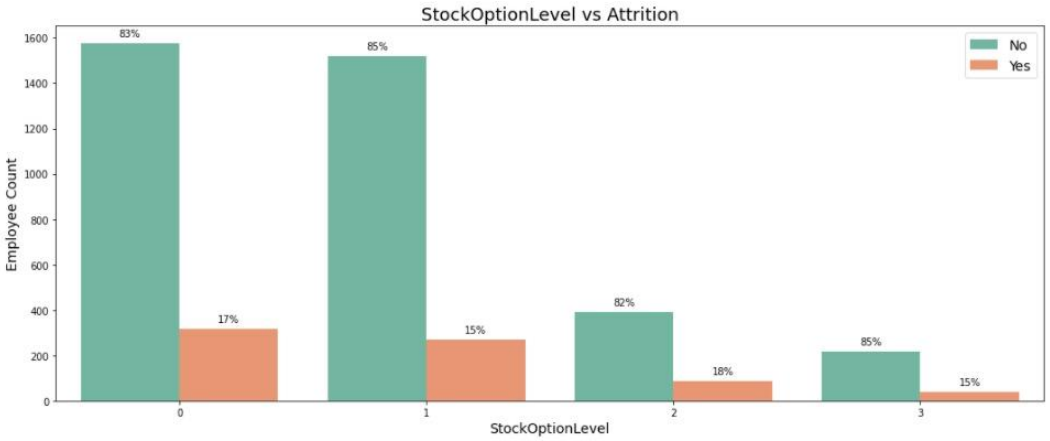
- Un-married Employees Have much higher attrition chance compared to married/divorced employees.



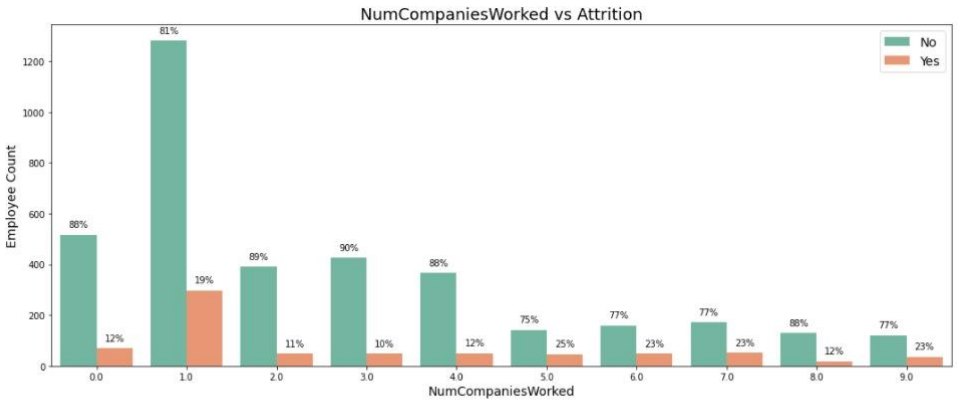
- Not much difference in attrition with respect to performance rating.



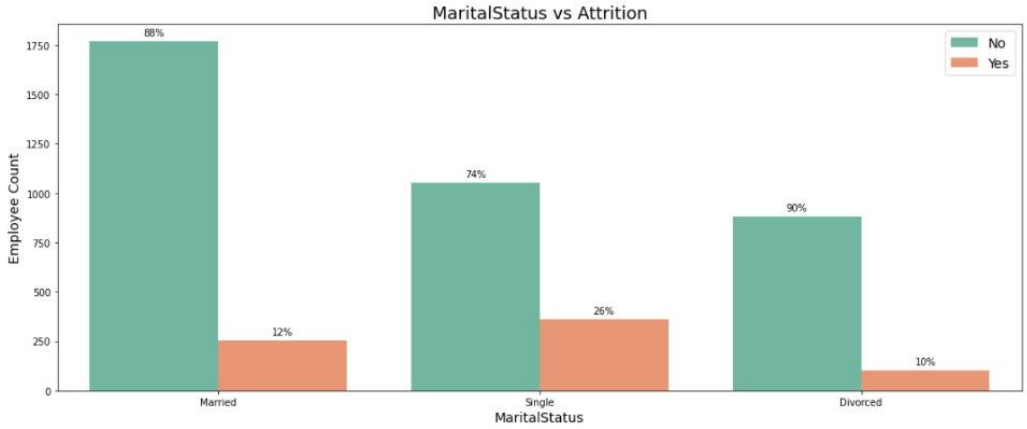
- Employee with Low job involvement, tend to have a higher Attrition %.



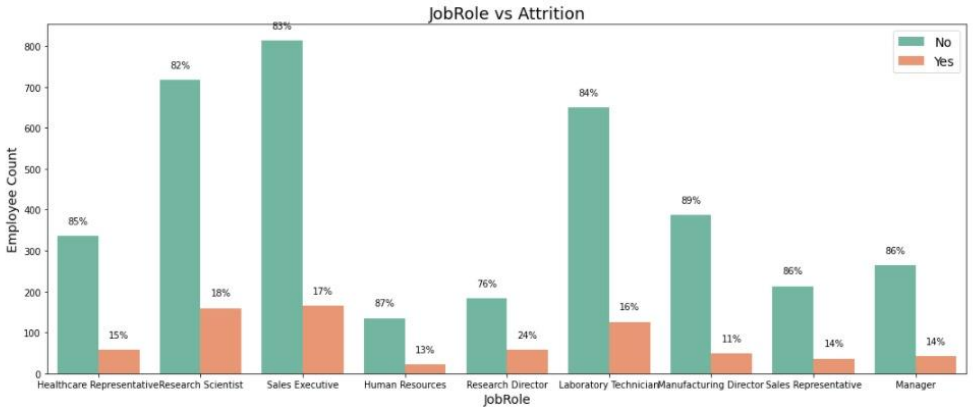
- Not much effect on Attrition with respect to employees StockOptionLevel.

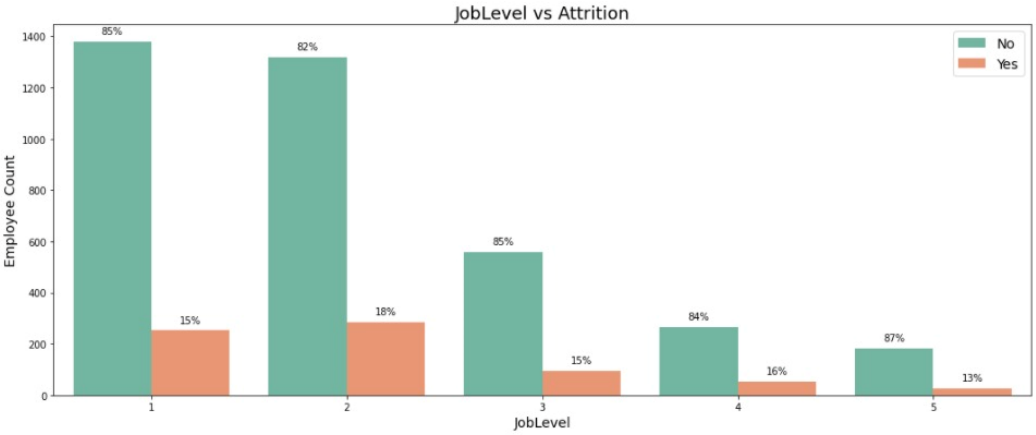


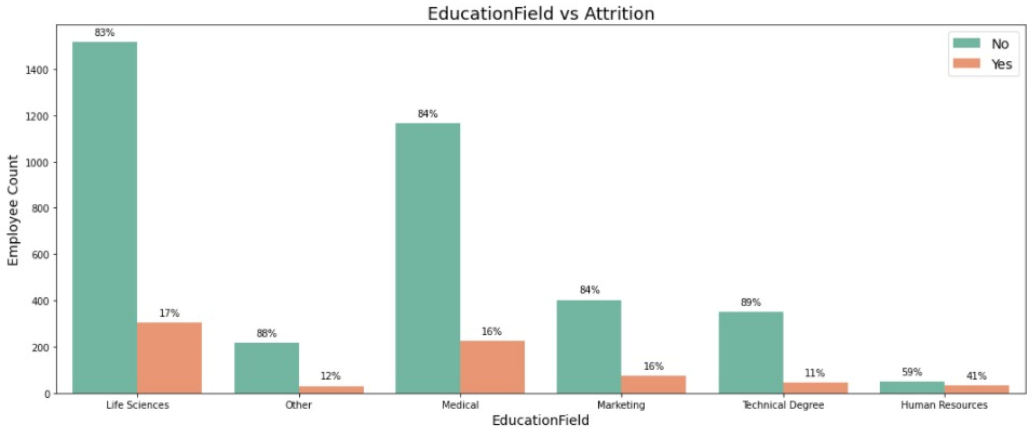
There is no clear trend in attrition with respect to number of companies an employees has worked. Altough employees who have worked in more than 4 companies are highly prone to leave the company.

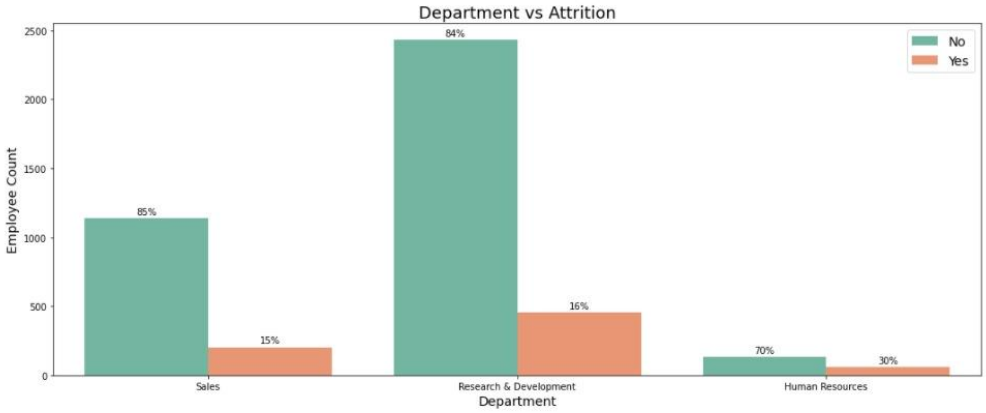


- Un-married Employees Have much higher attrition chance compared to married/divorced employees.

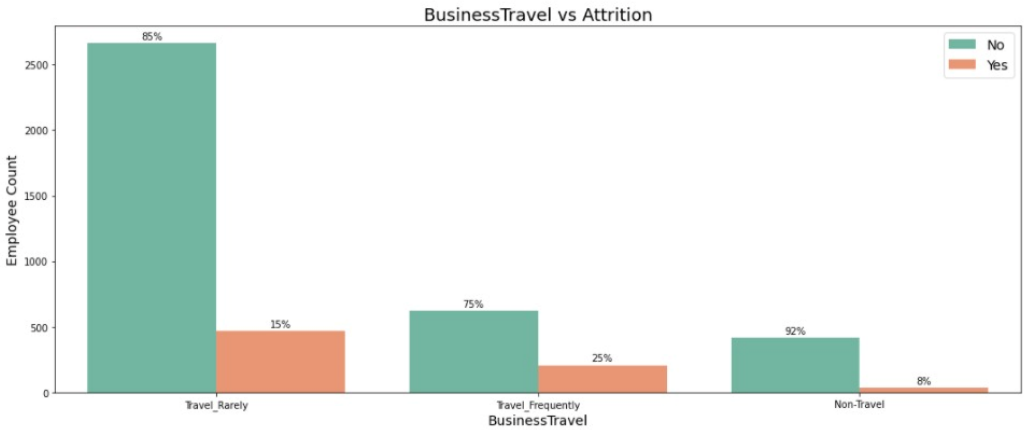








- Employees in HR Department have much Higher Attrition % compared to Employees in other Departments.



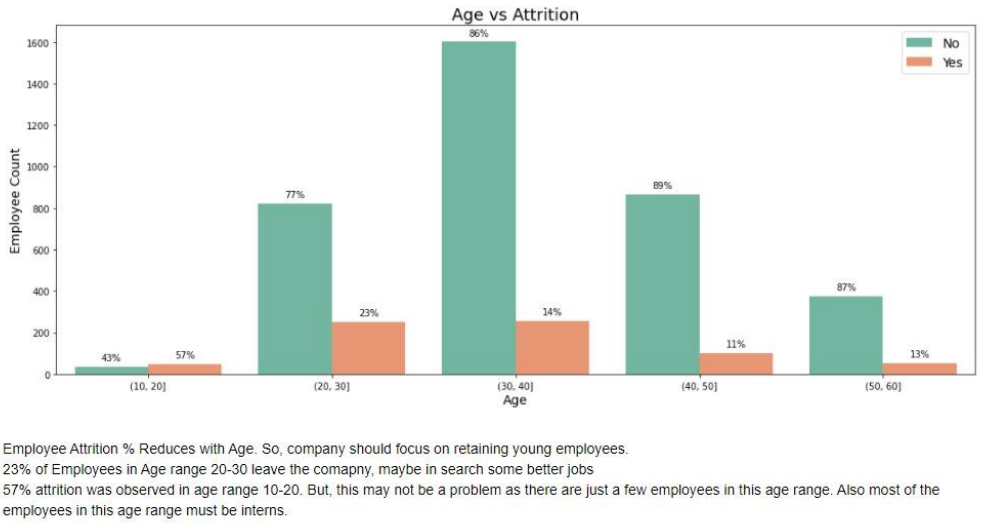


Figure 5.2 Data Visualizations of Dataset

The graph shown are: -

- Marital Status vs. Attrition
- Work Life balance vs. Attrition
- Job Satisfaction vs. Attrition
- Environment Satisfaction vs. Attrition
- Job Involvement vs. Attrition
- Education field vs. Attrition
- Non-Companies Worked vs. Attrition
- Department vs. Attrition
- Age vs. Attrition
- Business Travel vs. Attrition

When the dataset is analyzed, it gave a clear view about data and some points to keep for further process some notable points from above EDA are: -

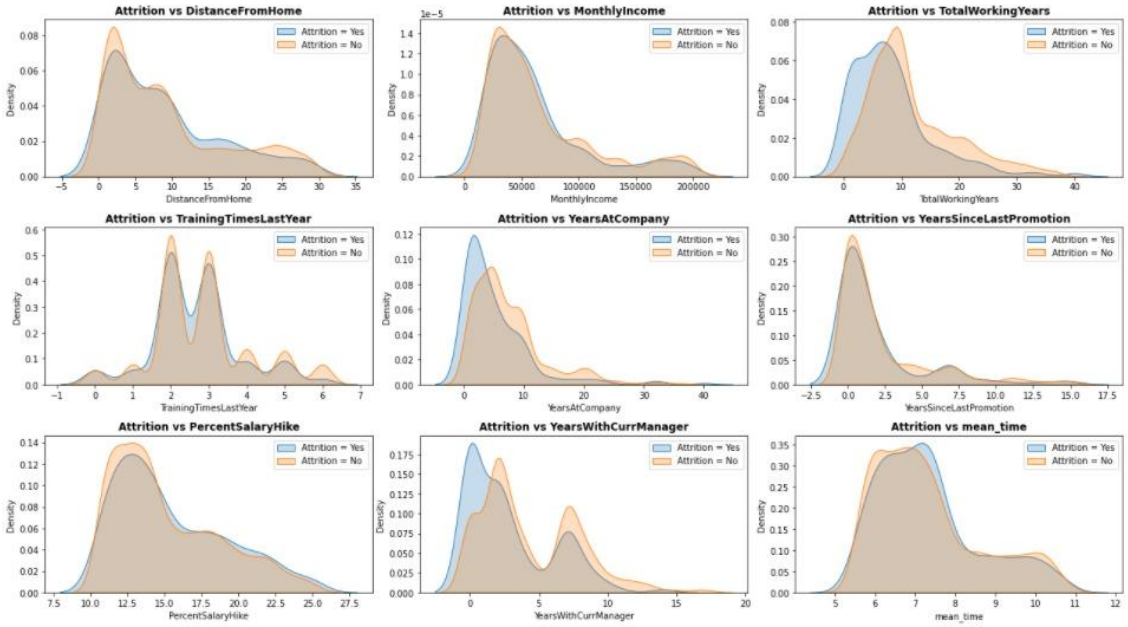
No trend seen between gender and rate of attrition similarly there was no concrete trace of any relation between previous experience, Education, and attrition.

All graphs point out that employees from HR department are leaving more.

Employees who have travel frequently are quitting more than others may be of more travels.

Attrition chances decreasing with age and can be added with unmarried people have more quitting ratio than married.

Environment, job, and work life balance are great factor for attrition response.



- From 'Attrition vs TotalWorkingYears' and 'Attritions vs YearsAtCompany' it is clear that, employees who stay at company for a longer period of time have lower attrition rate. Similar pattern is also observed in case of 'Attrition vs Years withCurrManager'.
- From 'Attrition vs mean_time' it can also be noted that the employees who left the company had slightly higher mean working hours. So. longer working Hours may be one of the reason for Attrition.

Figure 5.3 Density curve

In this division we are make use of exemplify the results of the experiments acted over the dataset [23]. We have judged the models established various versification like accuracy, recall, accuracy and f1 score.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad \text{.....[5.1]}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad \text{.....[5.2]}$$

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall} \quad \text{.....[5.3]}$$

SWOT Analysis

SWOT stands for Strength Weakness Opportunities SWOT can be defined as a structured way of evaluation. A method for evaluating a project in terms of its strengths weaknesses opportunities and threats involved now and can be arise later. It can also be referred as foundation for evaluating the overall potential and limitations of a project. It is also a feasibility test of the project. We try to include all potential risks and known limitations in it. It helps us to overcome the risks and improve the project quality.

SWOT analysis	HELPFUL (to achieve the objective)	HARMFUL (to achieve the objective)
Internal factors (attributes of the organization)	Strengths	Weaknesses
External factors (attributes of the environment)	Opportunities	Threats

Table 5.1 SWOT Representation

- **Strengths**

Huge Database of Detailed Reports of Past Projects

In companies all the information about commits and other software development related information are stored in an incredibly detailed reports manner. So, we can have a particularly good and big database of these reports by which we can extract our data that can be used for many purposes like to feed machine learning models etc.

Broad Range of Sample Projects

After completion of learning the model, we can also have a broad range of sample projects on which we can evaluate our model accuracy and other evaluation metrics. This will help us evaluating the best model and improving the prediction rate.

Assorted Studies Present

We can take reference from wide range of research papers, articles, journals, and other open-source material present related to this topic. There is a vast range of studies going to study the impact of AI in SDLC.

Selected models with results are present

On software fault prediction in a wide range of studies to you from which we can conclude and select the best models for learning purpose there are reasons and other limitations are also mentioned in those papers by which it will help us in understanding the reason of

higher accuracy and lower accuracy.

Experts for Feedback

The results from the eighth model are easily be understood by experts or anyone from this field so that they can cross validate those results and give us quality feedback for our learning and results.

- **Weaknesses**

  Very Broad Dataset

  The data set we have Stuart broad which will require a lot of data preprocessing and a lot of time for model to learning the complete data set.

  Complex Goals

  The girls are trying to achieve by artificial intelligence in software testing are too complex and have distinct types which will make the model complex and may need different models for different goals.

  Limited Approaches Defined

  In all the studies and journals, an extremely limited approach is used for learning and predicting the results the AI used in doors models is too narrow which narrow down the scope of the project.

  Irregular Dataset

  The data set we are going to have been most probably in Jira files or reports which are descriptive in nature so there will be a lot of data loss in converting those descriptive data sets in favorable nature

  Adaptability for New Type of projects

  The world is changing day by day as well as their needs show the projects going to come in future will have the need different from the past project and these an artificial intelligence thing pure only works on historical data so for those new type of projects this will not give a proper accuracy of predictions

- **Opportunities**

  Explore New Techniques

  All the research done in this field is very narrow in nature, so we can explore this and find new techniques to achieve the desired results

  Efficiency in the Whole Process

  More Goals Can be Added

The area and the goals can be e extended to more like in addition we can include many types of review system based on this model and we can also so expect the rating

Data Visualization Report for Developers

In the meantime, of learning we can create several data visualization report based on all data which will clear the view for them and help them in choosing the right testing or test cases

- **Threats**

Narrow Database

May the database present be broad, but it is possible that all the database is of same type of projects so if a project comes of slightly different requirements, then this model will fall to predict desired results.

Overfitting

This will be a great threat in using artificial intelligence in software testing as it is possible that there are only some types of testing or test cases going on so the air model will only have considered them and cannot add new to this

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

SWOT analysis can be used to conclude that the prediction for software faults and bugs depends a lot on dataset and one must rely on dataset a lot. it is a major drawback also because if anyone do not have much data or previous projects than making an AI model would be a difficult task. To prepare a model one must work more on extracting data. also, one should use the dedicated algorithms like SZZ for

bug finding and use SFP for fault prediction using previous work will save time as well as energy.

In employee turnover prediction xgboost proves as a better algorithm for prediction. Using good dataset and feature extraction helps in accuracy.

For future work we will use unique features to study more the problem of employee turnover and dependency of prediction over dataset.

References

[1] E. Ngai, L. Xiu, and D. Chau," Application of data mining techniques in customer relationship management: A literature review and classfication," Expert Systems with Applications, vol. 36, no. 2, Part 2, pp. 2592 {2602, 2009.

[2] S. Kaur and R. Vijay, "Job Satisfaction – A Major Factor Behind Attrition or Retention in Retail Industry," Imperial Journal of Interdisciplinary Research, vol. 2, no. 8, 2016.

[3] K.-B. Duan and S. S. Keerthi, "Which is the best multiclass SVM method? An empirical study," International workshop on multiple classifier systems, 2005.

[4] P. Cunningham and S. J. Delany, "k-Nearest neighbor classifiers," Multiple Classifier Systems, 1-17, 2007.

[5] Rish, Irina, "An empirical study of the naive bayes classifier," IJCAI Workshop on Empirical Methods in AI. C. Cortes and V. Vapnik, Support-vector networks. Machine learning, 20(3), 273-297, 1995.

[6] "A decade of agile methodologies: Towards explaining agile software development" Journal of Systems and Software, Volume 85, Issue 6, June 2012, Pages 1213-1221

[7] "A literature review of empirical research methodology in lean manufacturing"
Atmaca, E., Girenes, S.S. Lean Six Sigma methodology and application. Qual Quant 47, 2107–2127 (2013).

[8] Petersen K., Wohlin C., Baca D. (2009) The Waterfall Model in Large-Scale Development. In: Bomarius F., Oivo M., Jaring P., Abrahamsson P. (eds) Product-Focused Software Process Improvement. PROFES 2009. Lecture Notes in Business Information Processing, vol 32. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02152-7_29

[9] Saad Masood Butt, Wan Fatimah Wan Ahmad "The Waterfall Model in Large-Scale Development" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1, May 2012

[10] "An overview of Software Models with Regard to the Users Involvement"
Deepika Ganeshan, Published: 06 Jun 2011 "IBM Study on Software Models" Waterfall versus Agile methods: A pros and cons analysis""

[11] C. Ebert, G. Gallardo, J. Hernantes and N. Serrano, "DevOps," in IEEE Software, vol. 33, no. 3, pp. 94-100, May-June 2016, doi: 10.1109/MS.2016.68.

[12] Sonali Mathur, Shaily Malik. ©2010 International Journal of Computer Applications (0975 – 8887) Volume 1 – No. 12 "Advancements in the V-Model"

[13] BDD framework for .net - specflow - find bugs before they happen. BDD framework for NET. (2021, November 18). Retrieved May 26, 2022, from https://specflow.org/

[14] P. Ramya, V. Sindhura, and P. V. Sagar, "Testing using selenium web driver," 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2017, pp. 1-7, doi: 10.1109/ICECCT.2017.8117878.

[15] Harpreet Kaur, Dr. Gagan Gupta, "Comparative Study of Automated Testing Tools: Selenium, Quick Test Professional and Test complete" Harpreet kaur et al Int. Journal of Engineering Research and Applications www.ijera.com ISSN: 2248-9622, Vol. 3, Issue 5, Sep-Oct 2013, pp.1739-1743

[16] Ruchika Malhotra, "A systematic review of machine learning techniques for software fault prediction" Volume 27, February 2015, Pages 504-518.

[17] Cagatay Catala, Banu Dirib," A systematic review of software fault prediction studies" Volume 36, Issue 4, May 2009, Pages 7346-7354

[18] Rathore, S.S., Kumar, S. A study on software fault prediction techniques. Artif Intell Rev 51, 255–327 (2019). https://doi.org/10.1007/s10462-017-9563-5

[19] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," Journal of Machine, 2011.

[20] Mauricio A. Valle & Gonzalo A. Ruz (2015) Turnover Prediction in a Call Center: Behavioral Evidence of Loss Aversion using Random Forest and Naïve
Bayes Algorithms, Applied Artificial Intelligence, 29:9, 923-942, DOI: 10.1080/08839513.2015.1082282.

[21] International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)269 "Predicting Employee Attrition along with Identifying High Risk Employees using Big Data and Machine Learning " , Apurva Mhatre, Avantika Mahalingam, Mahadevan Narayanan, Akash Nair, Suyash Jaju.

[22] Lessmann, Stefan, and Stefan Voß. "A reference model for customercentric data mining with support vector machines." European Journal of Operational Research 199.2 (2009): 520-530.

[23] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters 27), 861–874, 2006.

[24] A New Multi-layer Perceptrons Trainer Based on Ant Lion Optimization Algorithm Raschka, S. (2015). Python machine learning. Packt Publishing Ltd.

[25] Morgan, J.N., Sonquist, J.A.: Problems in the analysis of survey data, and a proposal. J. Am. Stat. Assoc. 58, 415–434 (1963)

[26] Cross validation in R: Usage, Models & Measurement. upGrad blog. (2021, November 29). Retrieved

May 26, 2022, from https://www.upgrad.com/blog/cross-validation-in-r/

[27] Géron, A.: Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly Media (2017)

[28] Zhao, Yue, et al. "Employee turnover prediction with machine learning: A reliable approach." Proceedings of SAI intelligent systems conference. Springer, Cham, 2018.

[29] Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. 20, 273– 297 (1995)

[30] "How do I select SVM kernels?" Dr. Sebastian Raschka, May-2020. [Online]. Available: https://sebastianraschka.com/faq/docs/select_svm_kernel. html. [Accessed: 14-May-2020].

[31] Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794, ACM (2016).

[32] Singh, Jay P. "Predictive validity performance indicators in violence risk assessment: A methodological primer." Behavioral Sciences & the Law 31.1 (2013): 8-22.

[33] I. Žliobaite, "Learning under concept drift: an overview," Computing Research Repository by Cornell library, pp. –1–1, 2010.

[34] L. C. Briand, Y. Labiche, and Z. Bawar, "Using Machine Learning to Refine Black-Box Test Specifications and Test Suites," 2008 The Eighth International Conference on Quality Software, 2008.

[35] G. Grano, T. V. Titov, S. Panichella, and H. C. Gall, "How high will it be? Using machine learning models to predict branch coverage in automated testing," 2018 IEEE Workshop on Machine Learning Techniques for Software Quality Evaluation (MaLTeSQuE), 2018.

[36] A. Rauf and M. N. Alanazi, "Using artificial intelligence to automatically test GUI," 2014 9th International Conference on Computer Science & Education, 2014.

[37] D. J. Mala and V. Mohan, "IntelligenTester –Test Sequence Optimization Framework using Multi-Agents," Journal of Computers, vol. 3, no. 6, Jan. 2008.

[38] Y. Pang, X. Xue, and A. S. Namin, "Identifying Effective Test Cases through K-Means Clustering for Enhancing Regression Testing," 2013 12th International Conference on Machine Learning and Applications, 2013.

[39] T. Hall and D. Bowes, "The State of Machine Learning Methodology in Software Fault Prediction," 2012 11th International Conference on Machine Learning and Applications, 2012

[40] R. Gove and J. Faytong, "Identifying Infeasible GUI Test Cases Using Support Vector Machines and Induced Grammars," 2011 IEEE Fourth International Conference on Software

Testing, Verification and Validation Workshops, 2011.

[41] K. Chandra, G. Kapoor, R. Kohli, and A. Gupta, “Improving software quality using machine learning,” 2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH), 2016.

[42] R. Lachmann, S. Schulze, M. Nieke, C. Seidl, and I. Schaefer, “System-Level Test Case Prioritization Using Machine Learning,”2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), 2016.