

what is this project about?

regression tests

type of software testing to confirm that a recent program or code change has not adversely affected existing features.

automation

It's an automation of test behaviour.

Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Some machine learning methods

Machine learning algorithms are often categorized as supervised or unsupervised.

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training.

The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

In contrast, unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled.

Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data.

The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training – typically a small amount of labeled data and a large amount of unlabeled data.

The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabeled data generally doesn't require additional resources.

Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement

learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

Machine learning enables analysis of massive quantities of data. While it generally delivers faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. Combining machine learning with AI and cognitive technologies can make it even more effective in processing large volumes of information.

why this project is needed?

related to my work

existing system

we have RTC software which shows that which tests were ran, what was the last time taken by a particular test.

limitation is that

1) every single time the test runs,

It's time is not stored somewhere.

so every time we want to see the test timing we've to manipulate the DB.

2) In current system there is method to find out the average time taken by the test

3) so there is no mechanism that will provide to separate the test based on their execution behavior

Manual work

if we want to check for the test behaviour, we've to access the DB manually fire some queries and then we'll be able to predict the behaviour on that it is time taking process.

when testing the s/w

there are two type of tests

re-test

regression test

re-test are those test which are run to test the functionality or bug again to make sure that code is fixed.

reg-test are those when app is undergoes a code changes and to ensure that new code isn't affecting the behavior or functionality of the product.

while regression testing

we have to undergo following phases

RESET ALL(44k)

all test cases existing in the bucket or suit should be re-executed.

expensive, takes a lot of time and resources

reg-test selection:

instead of selecting all the test cases, some tests are selected from the entire suite.

Prioritization of tests

Depending on business impact, critical and frequent used functionality, tests are prioritized from the suite.

Curious about

AI

DS

ML

The **Jupyter Notebook** is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more.”

Matplotlib is a plotting library for the Python programming language.

UML – use case

The purpose of a use case diagram in UML is to demonstrate the different ways that a user might interact with a system.

Actor, system/scenario, goals

**Communication diagrams**, formerly known as collaboration diagrams, are almost identical to sequence diagrams in UML, but they focus more on the relationships of objects—how they associate and connect through messages in a sequence rather than interactions.

- **Rectangles** represent **objects** that make up the application.
- **Lines** between class instances represent the **relationships between** different parts of the application.
- **Arrows** represent the **messages** that are sent between objects.
- **Numbering** lets you know in **what order the messages are sent** and how many messages are required to finish a process.

they specifically focus on *lifelines*, or the processes and objects that live simultaneously, and the messages exchanged between them to perform a function before the lifeline ends.

Symbol Name	Description
Object symbol	Represents a class or object in UML. The object symbol demonstrates how an object will behave in the context of the system. Class attributes should not be listed in this shape.
Activation box	Represents the time needed for an object to complete a task. The longer the task will take, the longer the activation box becomes.
Actor symbol	Shows entities that interact with or are external to the system.
Package symbol	Used in UML 2.0 notation to contain interactive elements of the diagram. Also known as a frame, this rectangular shape has a small inner rectangle for labeling the diagram.
Lifeline symbol	Represents the passage of time as it extends downward. This dashed vertical line shows the sequential events that occur to an object during the charted process. Lifelines may begin with a labeled rectangle shape or an actor symbol.
Option loop symbol	Used to model if/then scenarios, i.e., a circumstance that will only occur under certain conditions.
Alternative symbol	Symbolizes a choice (that is usually mutually exclusive) between two or more message sequences. To represent alternatives, use the labeled rectangle shape with a dashed line inside.

## Common message symbols

Use the following arrows and message symbols to show how information is transmitted between objects. These symbols may reflect the start and execution of an operation or the sending and reception of a signal.

Symbol Name	Description
Synchronous message symbol	Represented by a solid line with a solid arrowhead. This symbol is used when a sender must wait for a response to a message before it continues. The diagram should show both the call and the reply.
Asynchronous message symbol	Represented by a solid line with a lined arrowhead. Asynchronous messages don't require a response before the sender continues. Only the call should be included in the diagram.
Asynchronous return message symbol	Represented by a dashed line with a lined arrowhead.
Asynchronous create message symbol	Represented by a dashed line with a lined arrowhead. This message creates a new object.
Reply message symbol	Represented by a dashed line with a lined arrowhead, these messages are replies to calls.
Delete message symbol	Represented by a solid line with a solid arrowhead, followed by an X. This message destroys an object.

Class diagrams are one of the most useful types of diagrams in UML as they clearly map out the structure of a particular system by modeling its classes, attributes, operations, and relationships between objects

- **Name:** The first row in a class shape.
- **Attributes:** The second row in a class shape. Each attribute of the class is displayed on a separate line.
- **Methods:** The third row in a class shape. Also known as operations, methods are displayed in list format with each operation on its own line.

Component diagrams are integral to building your software system.

they help your team understand the structure of existing systems and then build new ones.

Output-I  
124,383

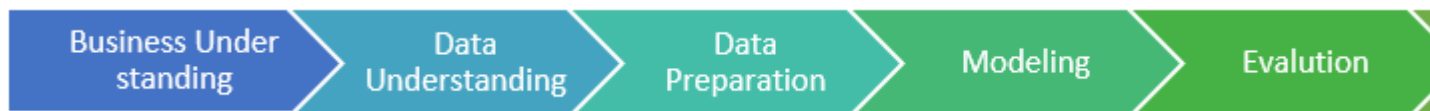
Retrieval test data

This is the data/test which are not used or performed more than 2 two times

Normal test data(48295)

Used/performed more than two times.

## Data Mining Implementation Process



Let's study the Data Mining implementation process in detail

### Business understanding:

In this phase, business and data-mining goals are established.

- First, you need to understand business and client objectives. You need to define what your client wants (which many times even they do not know themselves)

### Data understanding:

In this phase, sanity check on data is performed to check whether its appropriate for the data mining goals.

- First, data is collected from multiple data sources available in the organization.

### Data preparation:

In this phase, data is made production ready.

The data preparation process consumes about 90% of the time of the project.

The data from different sources should be selected, cleaned, transformed, formatted, anonymized, and constructed (if required).

Data cleaning is a process to "clean" the data by smoothing noisy data and filling in missing values.

## Data transformation:

Data transformation operations would contribute toward the success of the mining process.

## Modelling

In this phase, mathematical models are used to determine data patterns.

- Based on the business objectives, suitable modeling techniques should be selected for the prepared dataset.
- Create a scenario to test check the quality and validity of the model.
- Run the model on the prepared dataset.
- Results should be assessed by all stakeholders to make sure that model can meet data mining objectives.

**Standard deviation** is a number used to tell how measurements for a group are spread out from the average ([mean](#)), or expected value. A low standard deviation means that most of the numbers are close to the [average](#). A high standard deviation means that the numbers are more spread out.<sup>[1][2]</sup>

## Evaluation:

In this phase, patterns identified are evaluated against the business objectives.

## 2. Clustering:

Clustering analysis is a data mining technique to identify data that are like each other. This process helps to understand the differences and similarities between the data.