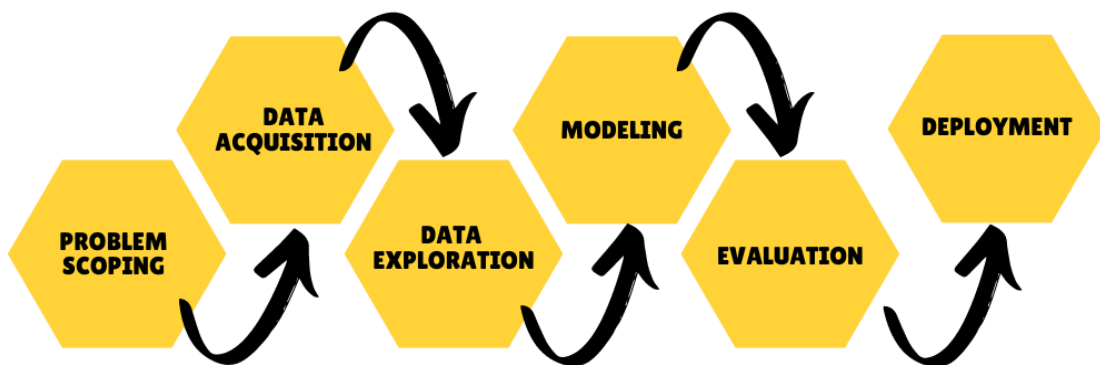# Unit 1.2: AI Project Cycle

In this chapter, you'll learn about the different stages of a project and how each stage contributes to the successful completion of a project. Understanding the project cycle is crucial for managing tasks effectively and ensuring that projects are completed on time and within budget.

Problem Scoping

Problem Scoping is the first stage of the AI project cycle. In this stage of AI development, problems will be identified. It is then followed by designing, developing, or building, and finally testing the project.

In the AI project cycle everything will fail if problem scoping is failed or without appropriate problem scoping. Incorrect problem scoping also leads to failure of the project as well.

The problem scoping refers to the identification of a problem and the vision to solve it.

The 4Ws of Problem Scoping

The 4Ws are very helpful in problem scoping. They are:

1. **Who?** - Refers that who is facing a problem and who are the stakeholders of the problem

2. **What?** - Refers to what is the problem and how you know

about the problem

3. **Where?** - It is related to the context or situation or location of the problem

4. **Why?** - Refers to why we need to solve the problem and what are the benefits to the stakeholders after solving the problem

Let us understand!

Let us go through the AI project cycle with the help of an example.

Problem: Pest infestation damages crops The cotton industry in India consists of 6 million local farmers. Cotton crops frequently get infected with the Pink Bollworm. It is difficult to see these insects with the naked eye. Small farmers find it very difficult to get rid of these insects. They do not have advanced tools and techniques to protect their plants from Pink Bollworm.
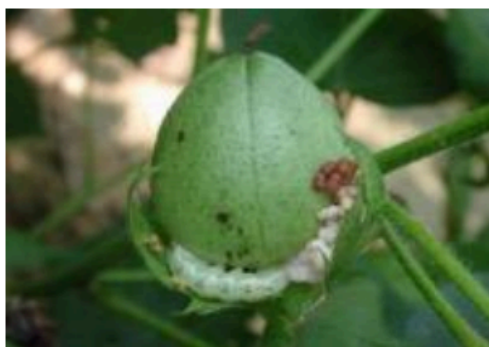
Can we solve this problem with AI? How?

Watch the video at this link - https://www.youtube.com/watch?v=LP_A4jydmz4

Start with listing down all the factors which you need to consider to save the cotton crop.

This system aims to:

_____

_____

__

While finalizing the aim of this system, you scope the problem which you wish to solve with the help of your project. This is **Problem Scoping.**

The problem statement template

When the above 4Ws are filled, you need to prepare a summary of these 4Ws. This summary is known as the problem statement template. This template explains all the key points in a single template. So, if the same problem arises in the future this statement helps to resolve it easily.

Our	[Stakeholders]	**who**
has a problem that	[issue, problem, need]	**what**
when/while	[context, situation]	**where**
An ideal solution would	[benefit of solution for them]	**why**

Now let us create a problem statement template for our Pest management case study

4W canvas for Pest Management

| Our	Farmers	**who**
has a problem that	Cotton Crops got infected with pest - Pink Bollworm	**what**
when/while	On the crops in the field	**where**
An ideal solution would	To create an AI-enabled app that aids farmers in- • Determining the correct amount of pesticides • Knowing the right time to spray pesticides	**why**

	● Increase in Production ● Increase in the profit share of the farmers.	

Activity - Brainstorm around the theme and set a goal for the AI project

In this activity, you need to select a theme for problem scoping.

Select the Theme

In CBSE Study Material, they have given the following themes for problem scoping:
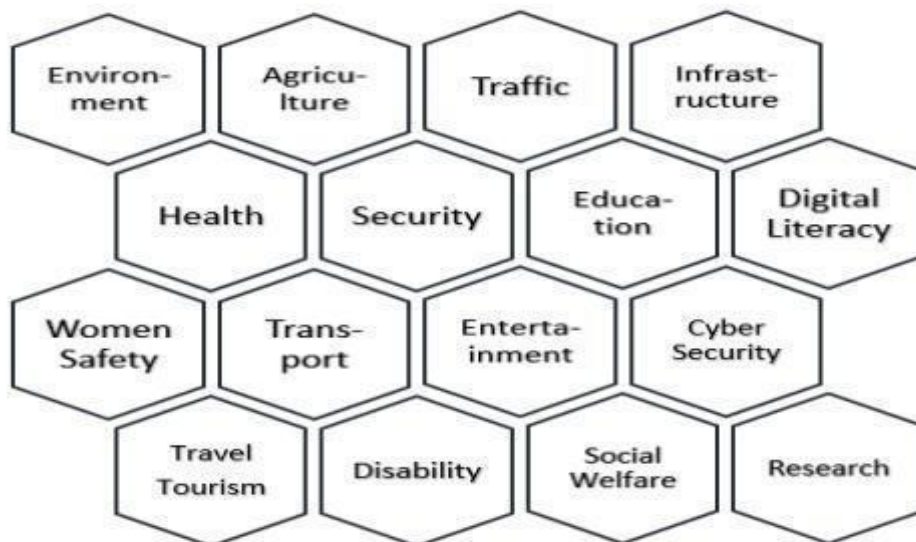
Reference CBSE Study Material

Students can select any of these or they can choose their own as well. For Example,

1. **The environment** is your theme. So think about the various

   problems such as polluted air, water, and land, etc.

2. Suppose you have selected **an Agriculture** theme, then there

   are various pesticides used in agriculture to increase the

   production, sowing and harvesting problems, etc.

3. Traffic is also one of the themes given in the handbook. Here

you can think about traffic issues and to reduce accidents or any other related problem.

Similarly you can take any theme and think about the various problems of that theme.

Place the problems into a problem statement template

After understanding and writing the problems, **set your goals**, and make them your AI project target. Write your goals for your selected theme.

Suppose you have selected the theme of agriculture then write how AI will help farmers to solve their problems.

1. Determine what will be a good time for seeding?

2. Determine what will be a good time for harvesting?

3. Determine when and how much fertilizer will be applied to the selected crop?

Your final problem statement will look like the following table for :

Who	Stakeholders
	Farmers, Fertilizer Producers, Labours, Tractor Companies
What	The problem, Issue, Need
	Determine what will a good time for seeding or crop harvesting?
When	Context/Situation
	Decide the mature age for the crop and determine its time
Ideal Solution	Benefits

	Take the crop on time and supply against market demand on time

Activity Time: Now pick a goal and prepare a problem template for your idea.

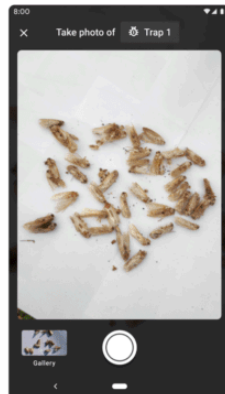|		

Let us look at the main features of CottonAce app-
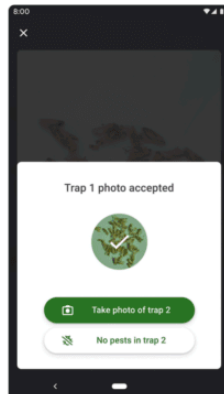

AI-based Pest Management App

A farmer or extension worker empties the moths from the pheromone trap.
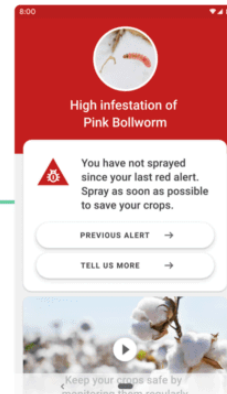
He then opens the app and captures 2 trap images.

The pests are detected and counted from the image uploaded.

An advisory is then generated on the app.

App Flow

WADHWANI AI

CottonAce app

- CottonAce is a mobile application that can help farmers protect their crops from pests.

- CottonAce uses AI to warn the farmers about a possible pest infestation.
- It aids farmers in –
  - Determining the correct amount of pesticides
  - Knowing the right time to spray pesticides

- Seeking professional help as needed.
- How does it work?
- A farmer sets up a trap to capture pests.
- Take a picture of the captured pests.
- Upload the picture on the app.
- The app detects the insect, level of infestation, and the required measures to cure it.

> You can add 'Small farms that used the app saw jumps in profit margins of up to 26.5%. A drop-in pesticide cost of up to 38 percent was also observed'.

Data Acquisition

Data Acquisition consists of two words:

1. **Data** : Data refers to the raw facts , figures, or piece of facts, or statistics collected for reference or analysis.

2. **Acquisition**: Acquisition refers to acquiring data for the project.

The stage of acquiring data from the relevant sources is known as data acquisition.

Classification of Data

Now Observe the following diagram to for the data classification, we will discuss each of them in detail:
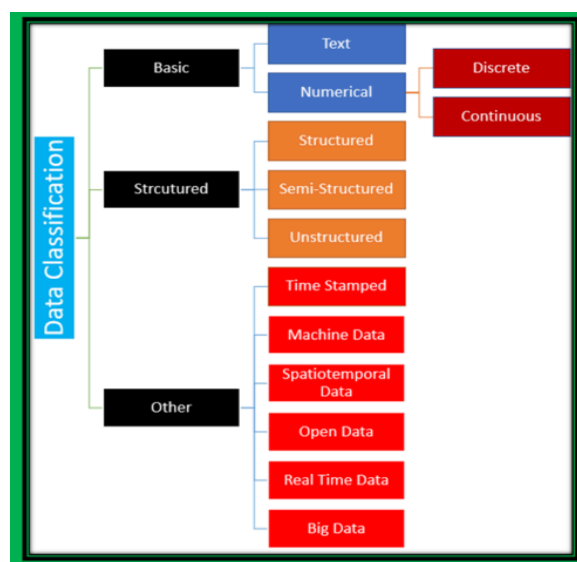
Classification of Data

Basic Data

Basically, data is classified into two categories:

1. **Numeric Data**: Mainly used for computation.
    o For example 132 customers, 126 Students etc.
    o For example 10.5 KGS, 100.50 Kms etc.
2. **Text Data**: mainly used to represent names, collection of words together, phrases, textual information etc.

Structural Classification

The data that is going to be feed in the system to train the model or already fed in the system can have a specific set of constraints or rules or a unique pattern can be considered as structural data.

The structure classification is divided into 3 categories:

1. **Structured Data**: As we discussed, structured data can have a specific pattern or set of rules. These data have a simple structure and store the data in specific forms such as tabular form. Example: The cricket scoreboard, Your school time table, Exam date sheet etc.

2. **Unstructured Data:** The data structure which doesn't have any specific pattern or constraints as well as can be stored in any form is known as unstructured data. Mostly the data that exists in the world is unstructured data. Example, Youtube Videos, Facebook Photos, Dashboard data of any reporting tool etc.

3. **Semi-Structured Data**: It is the combination of both structured and unstructured data. Some data can have a structure like a database whereas some data can have markers and tags to identify the structure of data.

Data Features

Data features refer to the type of data you want to collect. Here two terms are associated with this:

1. **Training Data**: The collected data through the system is known as training data. In other words the input given by the user in the system can be considered as training data.
2. **Testing Data:** The result data set or processed data is known as testing data. In other words, the output of the data is known as testing data.

Data Exploration

What is data exploration?

In the previous modules, you have set the goal of your project and have also found ways to acquire data. While acquiring data, you must have noticed that the data is a complex entity – it is full of numbers and if anyone wants to make some sense out of it, they have to work some patterns out of it. For example, if you go to the library and pick up a random book, you first try to go through its content quickly by turning pages and by reading the description before borrowing it for yourself, because it helps you in understanding if the book is appropriate to your needs and interests or not.

Thus, to analyze the data, you need to visualize it in some user-friendly format so that you can:

- Quickly get a sense of the trends, relationships and patterns contained within the data.
- Define strategy for which model to use at a later stage.
- Communicate the same to others effectively. To visualize data, we can use various types of visual representations.

Data Exploration refers to the techniques and tools used to visualize data through complex statistical methods.

Advantages of Data Visualization

• A better understanding of data • Provides insights into data • Allows user interaction • Provide real-time analysis • Help to make decisions • Reduces complexity of data	• Provides the relationships and patterns contained within data • Define a strategy for your data model • Provides an effective way of communication among users

Data Visualization Tools

There are many data visualization tools available like:
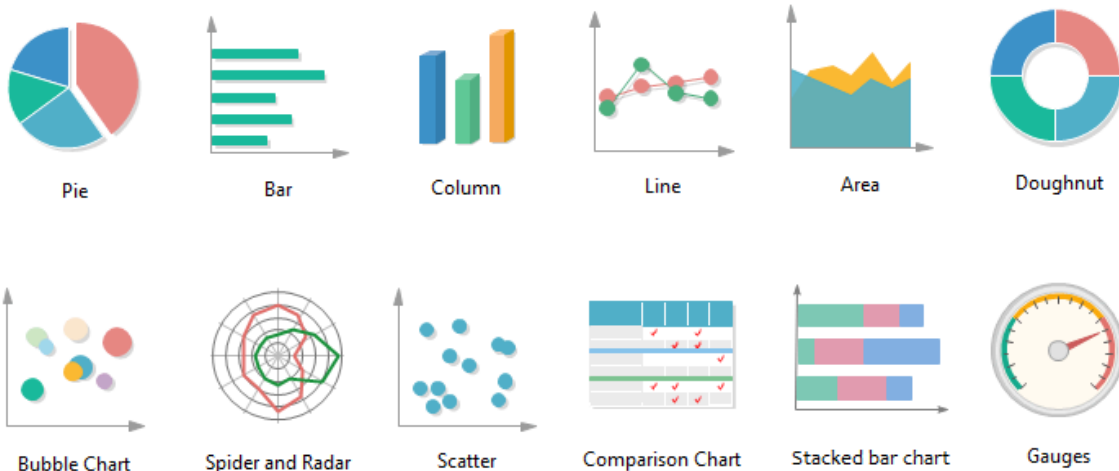
Let us now look at the scoped Problem statement and the data features identified for achieving the goal of your project. Try looking for the data required for your project from reliable and authentic resources. If you are not able to find data online, try using other methods of acquiring the data (as discussed in the Data Acquisition stage).

Once you have acquired the data, you need to visualize it. Under the sketchy graphs activity, you will visualize your collected data in a graphical format for better understanding.

For this, select one of the representations from the link or choose the ones which you already know. The basis of your selection should be the data feature which you want you to visualize in that particular representation. Do this for all the data features you have for the problem you have scoped. Let us answer the following questions for a better understanding:

1. Which data feature are you going to represent?

_____

_____

_____

_____

2. Which representation are you going to use for this data feature? Why?

_____

_____

_____

_____

Now, let's start drawing visual representations for all the Data features extracted, and try to find a pattern or a trend from it.

For example, if the problem statement is: How can we predict whether a song makes it to the billboard top 10?

We would require data features like: Current trends of music, genre of music, tempo of music, duration of song, popularity of a singer, etc.

Now to analyze a pattern, we can say that the popularity of the singer would directly have an effect on the output of the system. Thus, we would plot a graph showing the popularity of various singers and the one who is most popular has the maximum chance of getting to the billboard. In this way, the graphical representation helps us understand the trends and patterns out of the data collected and to design a strategy around them for achieving the goal of the project.

Modelling

In previous article data exploration, we have seen how we can represent data in graphics using various tools. This graphical representation makes data easy to understand for humans to make a decision or prediction. However, when it comes to machines to access and analyze data, machines require mathematical representation of data. Hence, every model needs a mathematical approach to analyze data. This could be done either by designing your own model or by using the pre-existing AI models. Before jumping into modelling let us clarify the definitions of Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL).

Defining the terms:

1. **Artificial Intelligence**, or AI, refers to any technique that enables computers to mimic human intelligence. The AI-enabled machines think algorithmically and execute what they have been asked for intelligently.

2. **Machine Learning,** or ML, enables machines to improve at tasks with experience. The machine learns from its mistakes and takes them into

consideration in the next execution. It improvises itself using its own experiences.

3. **Deep Learning**, or DL, enables software to train itself to perform tasks with vast amounts of data. In deep learning, the machine is trained with huge amounts of data which helps it into training itself around the data. Such machines are intelligent enough to develop algorithms for themselves.
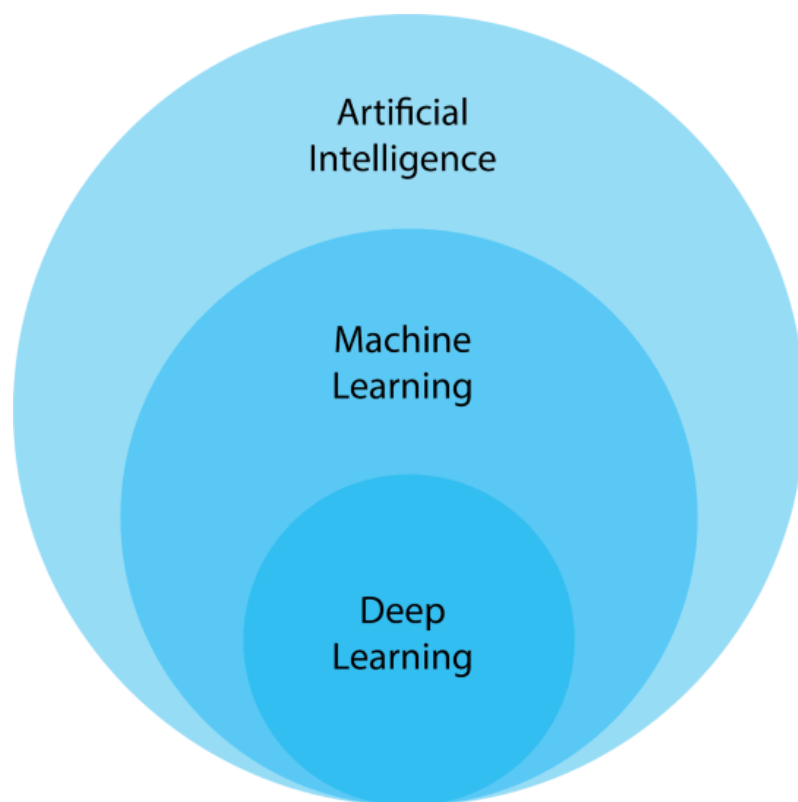
As you can see in the Venn Diagram, Artificial Intelligence is the umbrella terminology which covers machine and deep learning under it and Deep Learning comes under Machine Learning. It is a funnel type approach where there are a lot of applications of AI out of which few are those which come under ML out of which very few go into DL.

AI modelling approaches
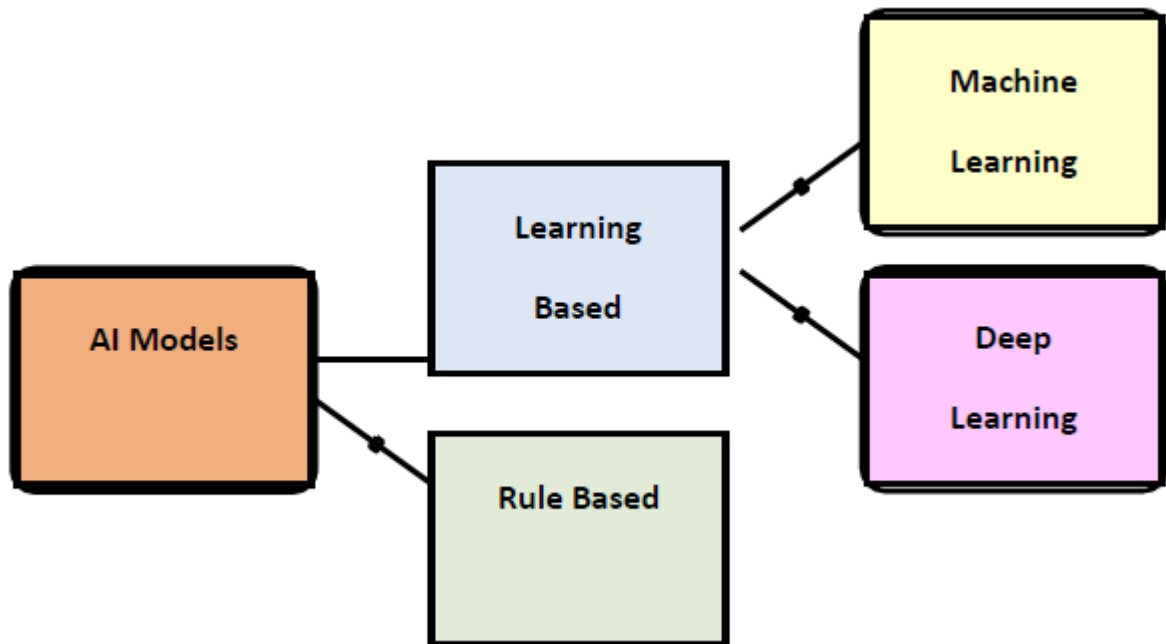
There are two approaches broadly taken by researchers for AI modelling. They are:

1. Rule-Based Approach
2. Learning-Based Approach

Let us begin with a rule-based approach.

1. Rule Based Approach

A Rule-based approach is generally based on the data and rules fed to the machine, where the machine reacts accordingly to deliver the desired output.

In other words, rule-based learning follows the relationship or patterns in data defined by the developer. The machine follows the instructions or rules mentioned by the developer and performs the tasks accordingly. It uses coding to make a successful model.

Consider the following scenarios and try to understand the rule-based approach for AI project Cycle modelling:

Suppose you have a dataset consisting of 100 images of apples and 100

images of bananas. To train your machine, you feed this data into the machine and label each image as either apple or banana.

Now if you test the machine with the image of an apple, it will compare the image with the trained data and according to the labels of trained images, it will identify the test image as an apple.

This is known as the Rule-based approach. The rules given to the machine in this example are the labels given to the machine for each image in the training dataset. Observe the following image:
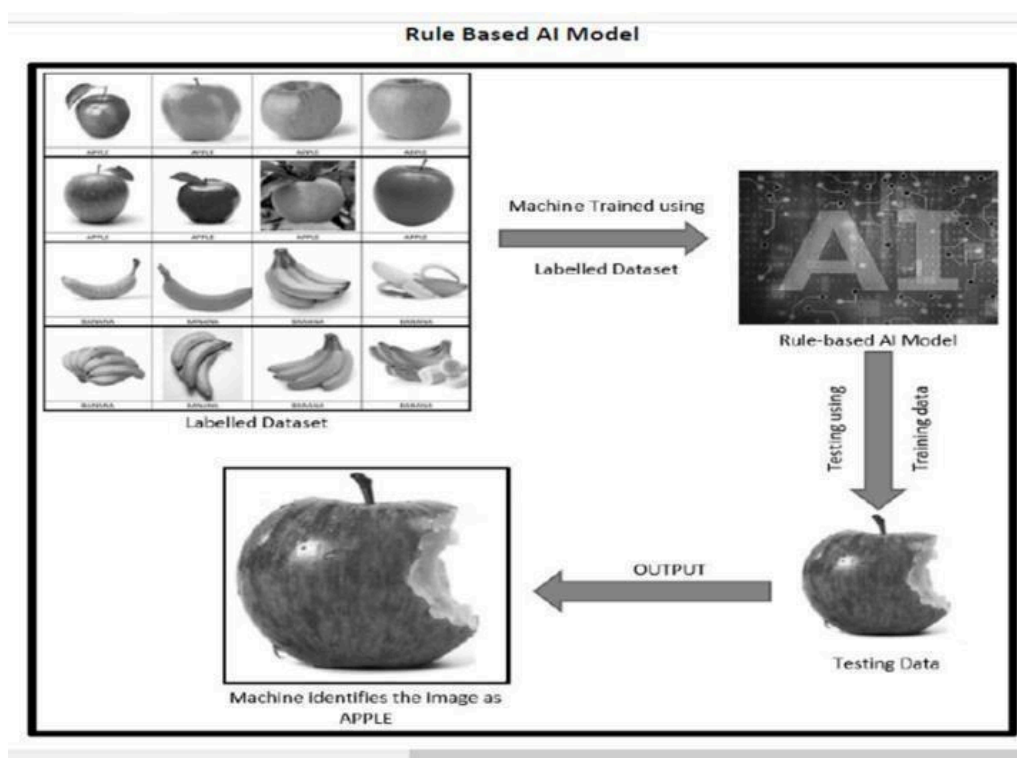
Decision Tree

The decision tree is one of the most common and basic models in data science. It follows a treelike structure of the decisions with all possible results.

The decision tree is made up of various nodes. It follows a top to bottom approach. The top most node of the decision tree is known as **root.**

Then it continues down to the terminal node or leaf node. Arrow lines connect all these nodes with each other.

Activity 1 - Make a decision tree

Decision tree is an example of a rule-based approach.

The structure of decision starts with th**e root node** and ends with **leaves by connecting branches** having different conditions.

So following things you have to keep in mind before making the decision tree:

1. Observe your data carefully.
2. Decide what (data) will be your root
3. Decide what (data) will be your leaves
4. Now analyse the data properly and find out some unnecessary data tree In the above picture, one decision tree is given.

Class 9 AI How to make a decision

Activity 1:

Here the decision is something which is related to our daily activity.
The first condition is whether you are hungry or not? If yes then if you have $25 you can go to the restaurant and if not then you can buy a burger or if you are not hungry then you can go to sleep.

So the top question or condition here is - **Am I Hungry?** will be considered as

root and Yes or no will be branches.

The final decisions like going to sleep, going to a restaurant and buying a burger are leaves. Have I $25 for the interior node.

Based on this decision tree two questions are given in the curriculum handbook and here I will provide you the answers:

1. **How many branches does the tree shown above have? - 2**
2. **How many leaves does the tree shown above have? - 3**

Activity 2

The following is a dataset consisting of 4 parameters which lead to the prediction of whether an Elephant would be spotted or not.

The parameters, which affect the predictions, are Outlook, Temperature, Humidity and Wind.

Draw a Decision Tree for this dataset.

Outlook	Temperature	Humidity	Wind	Elephant Spotted?
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Common decisions are as following:

- If outlook = Sunny and Humidity = High, then Elephant Spotted = No
- If outlook = Sunny and Humidity = Normal, then Elephant Spotted = Yes
- If outlook = overcast, then Elephant Spotted = Yes
- If outlook = Rain and wind= Strong, then Elephant Spotted = No
- If outlook = Rain and wind = weak, then Elephant Spotted = Yes

Similarly you can prepare any decision tree according to the given dataset.
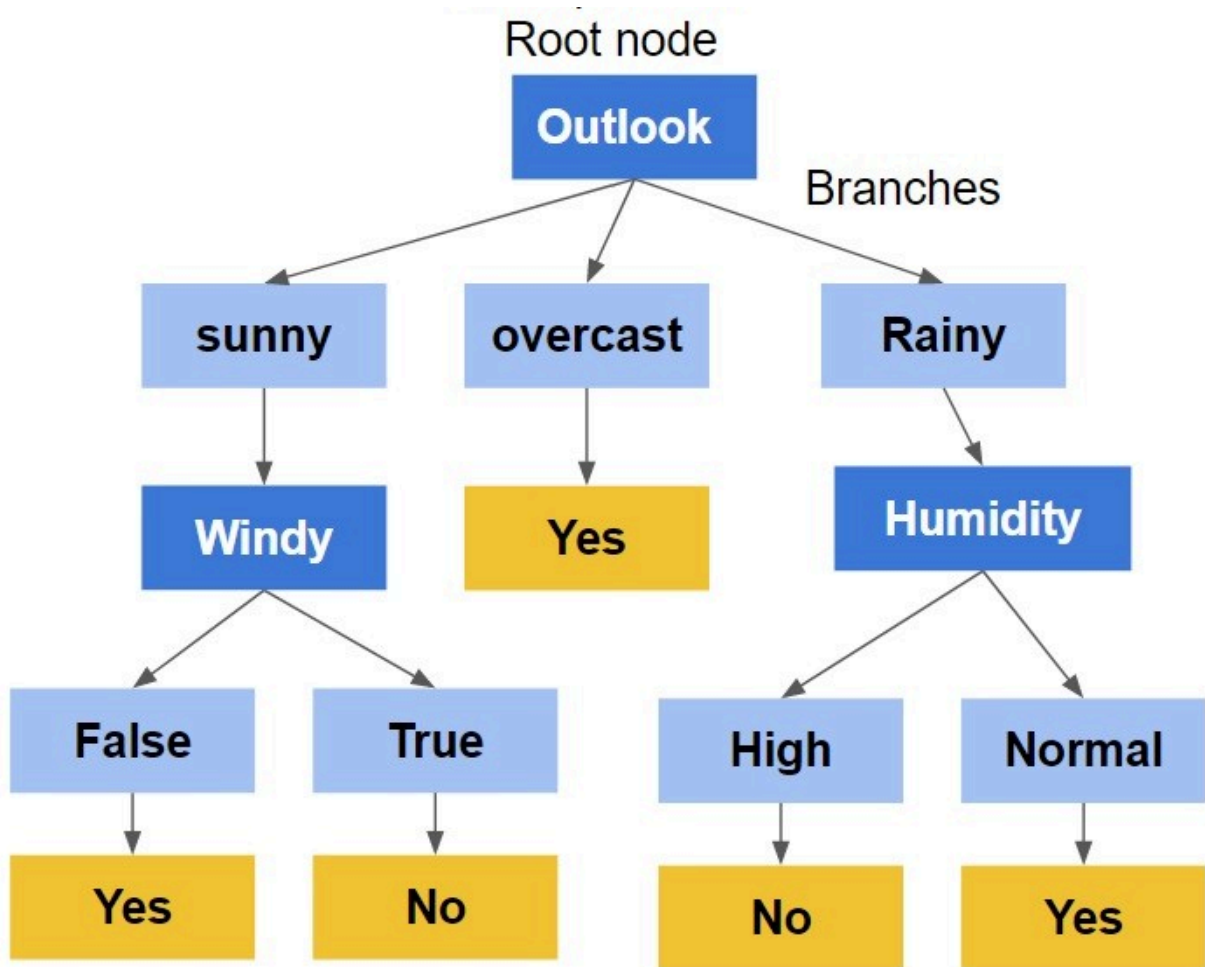
If you want to use an online websites to draw decision trees you can use one of the following online apps:

1. Creately : https://creately.com/lp/decision-tree-maker-online/
2. Smart Draw : https://www.smartdraw.com/decision-tree/decision-tree-maker.htm

3. Lucid Chart :
   https://www.lucidchart.com/pages/examples/decision-tree-maker
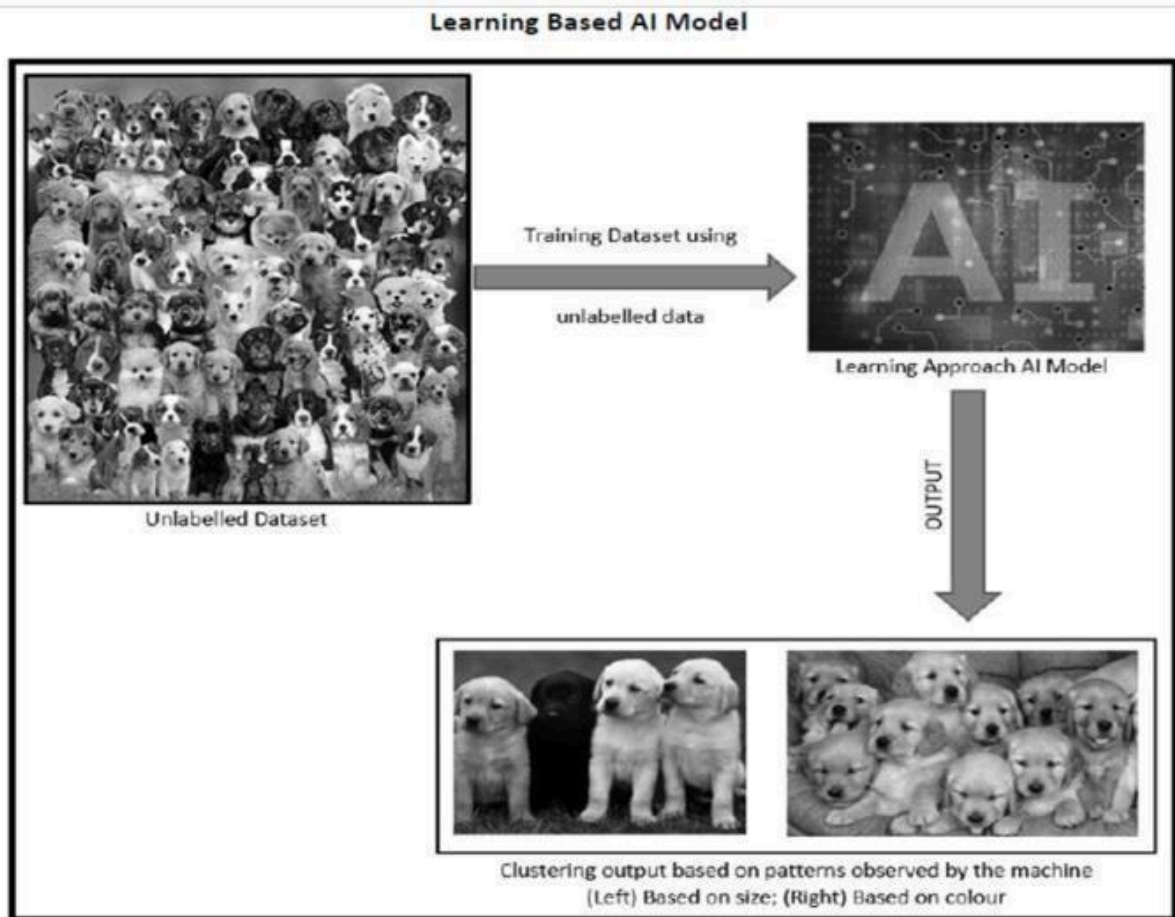
2. Learning Based Approach

The machine is fed with data and the desired output to which the machine designs its own algorithm (or set of rules) to match the data to the desired output fed into the machine to train.

In the learning-based approach, the relationship or pattern in data is not defined by the developer. This approach takes random data which is fed into the machine and it is left to the machine to figure out the patterns or required trends. In general this approach is useful when the data is not labelled and random for a human to use them.

Thus, the machine looks at the data, tries to extract similar features out of it and clusters the same datasets together. In the end as output, the machine tells us about the trends which are observed in the training data.

This approach is used to train the data, which is unpredictable, or the users have no idea about it. For example, suppose you have a dataset of 1000 images of random stray dogs of your area.Now you do not have any clue as to what trend is being followed in this dataset as you don't know their breed, colour or any other feature.

Thus, you would put this into a learning approach based AI machine and the machine would come up with various patterns it has observed in the features of these 1000 images. It might cluster the data on the basis of colour, size, fur style, etc. It might also come up with some very unusual clustering algorithm which you might not have even thought of!

Learning Based AI Model

Learning based approach - AI project cycle modelling

Evaluation

Evaluation is the process of understanding the reliability of any AI model, based on outputs by feeding a test dataset into the model and comparing with actual answers. There can be different Evaluation techniques, depending on the type and purpose of the model. Remember that It's not recommended to use the data we used to build the model to evaluate it. This is because our model will simply remember the whole training set, and will therefore always predict the correct label for any point in the training set. This is known as overfitting.

Once a model has been made and trained, it needs to go through proper testing so that one can calculate the efficiency and performance of the model. Hence, the model is tested with the help of Testing Data (which was separated out of the acquired dataset at Data Acquisition stage) and the efficiency of the

model is calculated on the basis of the parameters mentioned below:

Note: You will learn more about these techniques in grade X.

- We test our models to check their performance and improve our models for best performance.

- The model is tested with collected data.

- We also check if the model is solving the identified AI problem properly.

## Model Evaluation Terminologies

There are various new terminologies which come into the picture when we work on evaluating our model. Let's explore them with an example of the Autonomous Vehicle Pedestrian Detection.

The Scenario

Imagine that you have developed an AI-based prediction model for an autonomous vehicle to detect pedestrians. The objective of the model is to predict whether a pedestrian is present in the vehicle's path or not. To evaluate the efficiency of this model, we need to check if the predictions made by the model are correct or not. As with previous examples, there are two conditions to consider: Prediction and Reality. The prediction is the output provided by the model, and the reality is the actual presence or absence of a pedestrian.

Possible Outcomes

Case 1: The model predicts a pedestrian is present, and there is indeed a pedestrian.
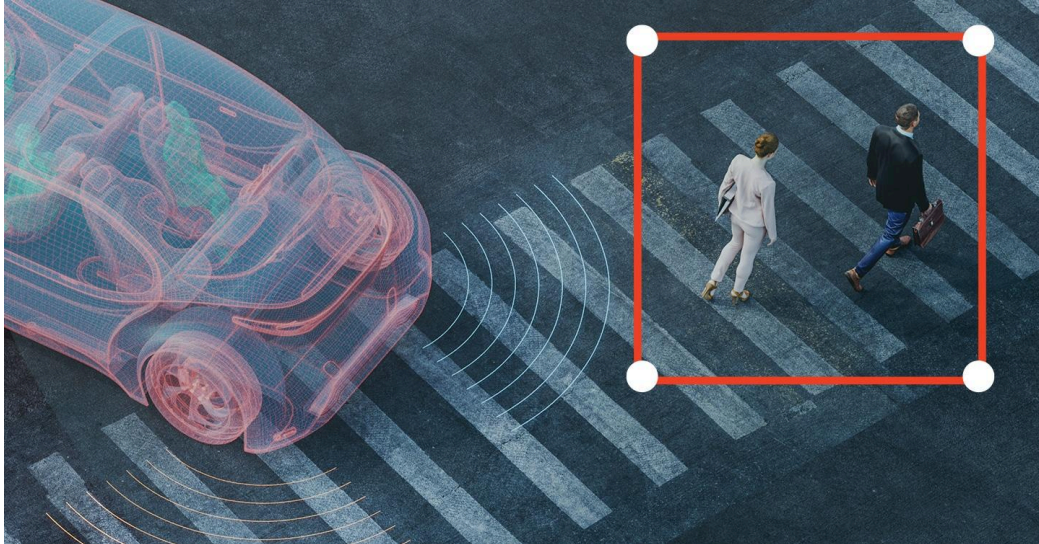
Prediction : Yes   Reality : Yes

True Positive

Case 2 : The model predicts no pedestrian is present, and there is indeed no pedestrian.

Prediction : No   Reality : No

True Negative

Case 3: The model predicts a pedestrian is present, but there is no pedestrian (Type I error)
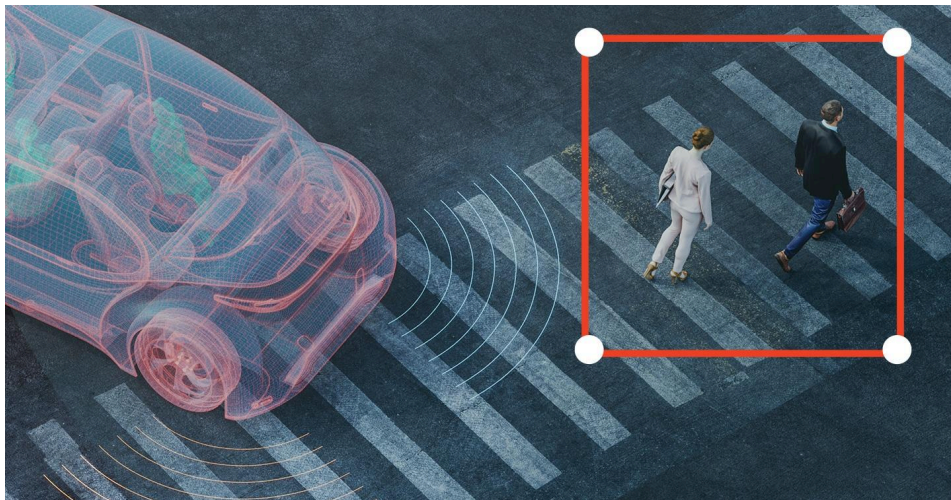

Prediction : Yes   Reality : No
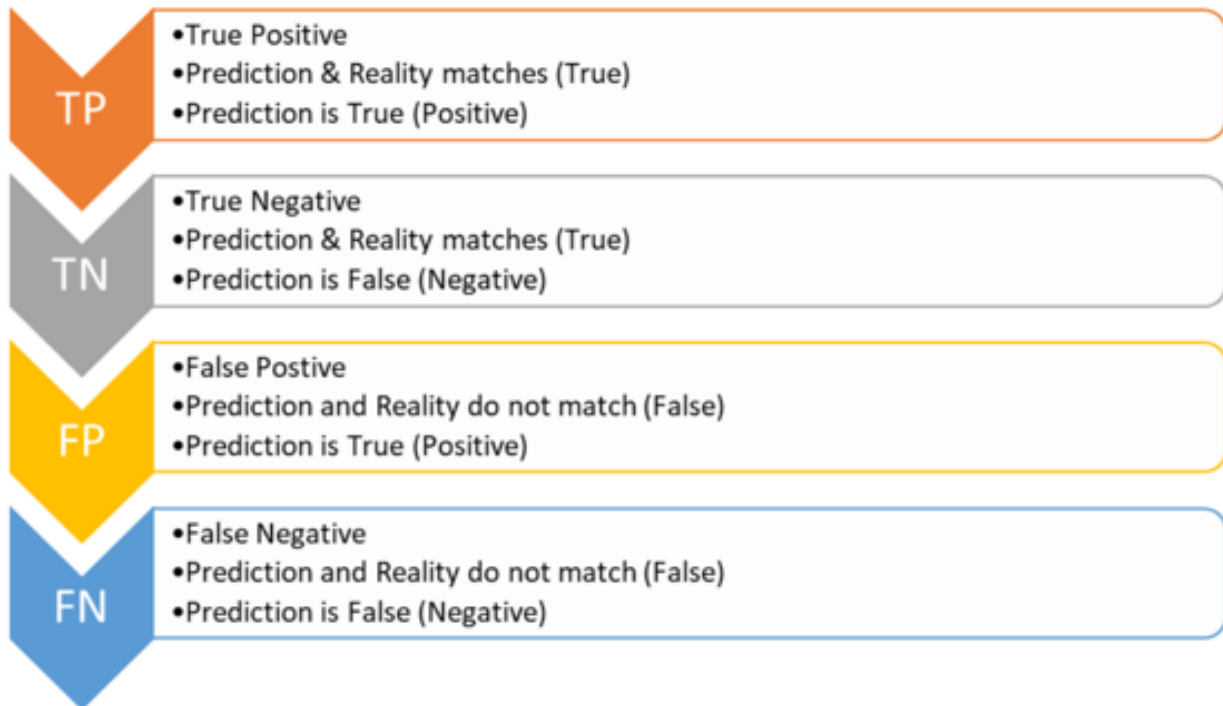
False Positive

Case 4 : The model predicts no pedestrian is present, but there is actually a pedestrian (Type II error).


Prediction : No   Reality : Yes

False Negative

TP
- True Positive
- Prediction & Reality matches (True)
- Prediction is True (Positive)

TN
- True Negative
- Prediction & Reality matches (True)
- Prediction is False (Negative)

FP
- False Postive
- Prediction and Reality do not match (False)
- Prediction is True (Positive)

FN
- False Negative
- Prediction and Reality do not match (False)
- Prediction is False (Negative)

The ROC curve is the plot of the true positive rate (TPR) against the false positive rate (FPR) at each threshold setting.

ROC is a metric used to find out the accuracy of a model.
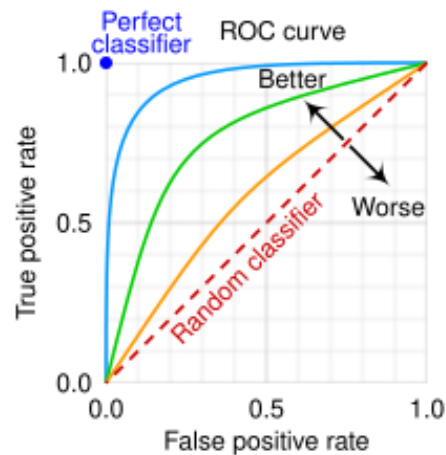
- We test our models to check their performance and improve our models for best performance.
- The model is tested with collected data.
- We also check if the model is solving the identified AI problem properly.

Deployment

Deployment as the final stage in the AI project cycle where the AI model or solution is implemented in a real-world scenario.

**Key Steps in Deployment :**

- Testing and validation of the AI model

- Integration of the model with existing systems

- Monitoring and maintenance of the deployed model.

Some examples of successful AI projects that have been deployed in various industries, such as self-driving cars, medical diagnosis systems, and chatbots.

Case Study: Preventable Blindness

Problem: Prevent loss of vision, and delay in report generation

- Approximately 537 million adults (20-79 years) are living with diabetes.
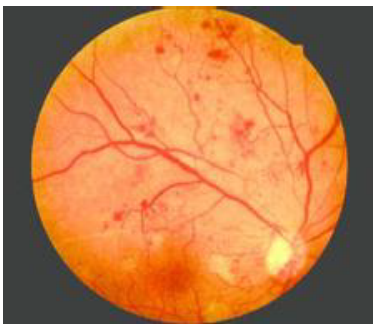
- Diabetes can lead to Diabetic Retinopathy It damages the blood vessels of the retina and can lead to blurred vision and blindness.

- Lack of qualified doctors and delay in reports increase the risk of Diabetic Retinopathy

One of the early symptoms of the defect is 'Blurred vision' as shown below:

Normal Vision   Vision with diabetic retinopathy

How can we solve this problem with AI?

Solution: **Using AI to detect Diabetic Retinopathy in pictures of eyes**

AI solution at Aravind Eye Hospital, India

A technician screening a patient at the Aravind Eye Hospital in Madurai, India. The hospital is using a Google system that relies on artificial intelligence to diagnose a retinal problem from such a scan. Credit...Atul Loke for The New York Times
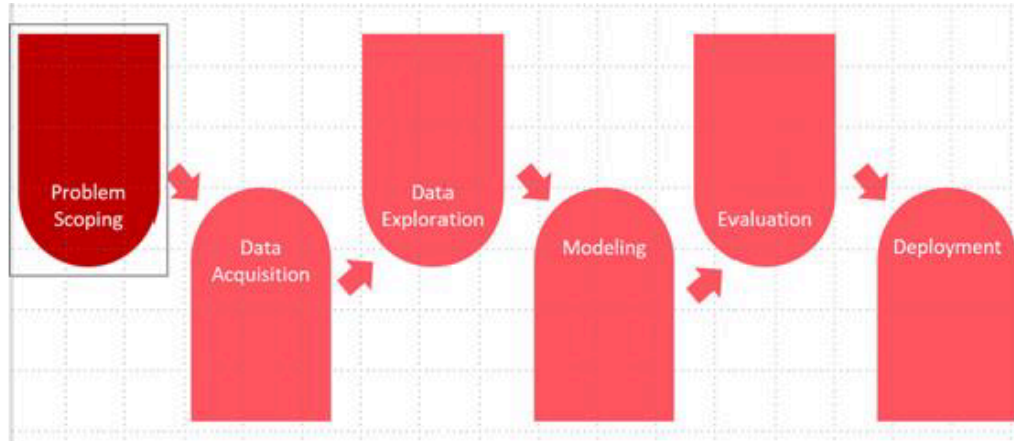
- An AI eye screening solution is developed in partnership with Google.

- AI models have achieved an accuracy of 98.6% in detecting diabetic retinopathy, on par with the performance of specialist eye doctors.

- Seventy-one vision centers in rural Tamil Nadu, India are using this solution.

- Trained technicians take pictures of patients' eyes with cameras.

- The digital images are analyzed by AI for the presence of Diabetic Retinopathy.

- AI has made the detection of Diabetic Retinopathy quicker.

- Any technician can use this machine, even without a skilled doctor.

More and more parents can be treated at an early stage.

Hence, early detection using AI can significantly benefit rural populations.

Let us map this problem to AI project cycle

How would you scope the problem?

AI Project Cycle Mapping Template				
Data Acquisition	**Data Exploration**	**Modeling**	**Evaluation**	**Deployment**
Collecting data from patients from many clinics using retinal cameras.	Validating all the data to make sense out of it and come up with a model.	Creating an AI model to correctly diagnose Diabetic Retinopathy when given a retinal image as input.	Test the model for accuracy and then fine tune the model further to get the desired out -put.	Using the model in tools that can be used in clinics in even the remote and rural parts of the country.

Activity Time!

Purpose: Implementation of AI project cycle to develop an AI Model

Suggested Topics:

- Personalized Education Plan
- Smart Traffic Management
- Personalized Fitness
- Smart Home Management
- Wildlife Conservation using Drones

Activity Introduction:

- In this activity, students use the AI project cycle to conceptualise a solution for the given problem.

- AI project cycle is a 6-step process which aids in problem solving using Artificial Intelligence

Description:

- All individuals have different cognitive levels and personalities.
- Different people need attention towards different parts of their learning.
- A generalized education system does not provide that.

Activity Guidelines:

- Understand the problem.
- Learn the various aspects and developments in the field.
- Fill the AI Project Cycle mapping template for the problem.
- The solution to the problem of personalized education is an AI algorithm that trains over the behavior and choices of a student. Thus, all the requirements specific to a student could be recognized and addressed to.

Fill the AI Project Cycle mapping template for the selected problem:

AI Project Cycle Mapping Template					
Problem Solving	Data Acquisition	Data Exploration	Modelling	Evaluation	Deployment

Exercise:

1. What are the various stages of the AI Project Cycle? Can you explain each briefly.
2. Why is there a need to use a Problem Statement Template during problem scoping?
3. How will you differentiate between Training Data and Testing Data? Elaborate with examples.
4. Name various methods for collecting data. For each method, can you name at least one project in which you may use that method of data

collection?

5. Name a few government websites from where you can get open-source data.

6. What do you think is the relevance of Data Visualization in AI?

7. List any five graphs used for data visualization.

8. How is Data Exploration different from Data Acquisition?

9. How is Machine Learning related to Artificial Intelligence?

10. Compare and contrast Rule-based and Learning-based approach in Al modeling indicating clearly when each of these may be used.

11. Identify which of the following are examples of classification/regression/clustering.

    I. Making a diagnosis for a patient on the basis of their symptoms

    II. Price prediction for a house coming up on sale

    III. HR shortlisting applications for interview based on information provided in candidates' resume

    IV. Credit Card Fraud prevention

    V. SPAM filters

12. What do you understand by the terms True Positive and False Positive?