K.T.S.P.Madal's

Hutatma Rajguru Mahavidyalaya, Rajgurunagar

Tal-Khed, Dist.-Pune 410505.

TY.BSc (Computer Science)

Semester-VI

Subject- Data Analytics

According to new CBCS syllabus w.e.f.2019-2020

Prof.S.V.Patole

Department of Computer Science

Hutatma Rajguru Mahavidyalaya,

Rajgurunagar.

Chapter 2.Machine Learning Overview

1.Introduction to Machine Learning:

A machine learning model is a program that can find patterns or make decisions from a previously unseen dataset. For example, in natural language processing, machine learning models can parse and correctly recognize the intent behind previously unheard sentences or combinations of words. In image recognition, a machine learning model can be taught to recognize objects - such as cars or dogs. A machine learning model can perform such tasks by having it 'trained' with a large dataset. During training, the machine learning algorithm is optimized to find certain patterns or outputs from the dataset, depending on the task. The output of this process - often a computer program with specific rules and data structures - is called a machine learning model.

Need of machine learning

- *Machine learning is important because it gives enterprises a view of trends in customer behavior and operational business patterns, as well as supports the development of new products.* Many of today's leading companies, such as **Facebook, Google,** and **Uber**, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.

- Machine learning has several practical applications that drive the kind of real business results - such as time and money savings - that have the potential to dramatically impact the future of your organization. In particular, we see tremendous impact occurring within the customer care industry, whereby machine learning is allowing people to get things done more quickly and efficiently.

- Through Virtual Assistant solutions, machine learning automates tasks that would otherwise need to be performed by a live agent - such as changing a password or checking an account balance. This frees up valuable agent time that can be used to focus on the kind of customer care that humans perform best: high touch, complicated decision-making that is not as easily handled by a machine.

Advantages of Machine Learning

1. Automation

Machine Learning is one of the **driving forces** behind automation, and it is cutting down time and human workload. Automation can now be seen everywhere, and the complex algorithm does the hard work for the user. Automation is more reliable, efficient, and quick. With the help of machine learning, now advanced computers are being designed. Now this advanced computer can handle several machine-learning models and complex algorithms. However, automation is spreading faster in the industry but, a lot of research and innovation are required in this field.

2. Scope of Improvement

Machine Learning is a field where things keep evolving. It gives many opportunities for improvement and can become the leading technology in the future. A lot of research and innovation is happening in this technology, which helps improve software and hardware.

3. Enhanced Experience in Online Shopping and Quality Education

Machine Learning is going to be used in the education sector extensively, and it will be going to enhance the quality of education and student experience. It has emerged in China; machine learning has improved student focus. In the e-commerce field, Machine Learning studies your search **feed and give suggestion** based on them. Depending upon search and browsing history, it pushes targeted advertisements and notifications to users.

4. Wide Range of Applicability

This technology has a very wide range of applications. Machine learning plays a role in almost every field, **like hospitality, ed-tech, medicine, science, banking, and business**. It creates more opportunities.

Disadvantages of the Machine Learning

Nothing is perfect in the world. Machine Learning has some serious limitations, which are bigger than human errors.

1. Data Acquisition

The whole concept of machine learning is about identifying useful data. The outcome will be incorrect if a credible data source is not provided. The quality of the data is also significant. If the user or institution needs more quality data, wait for it. It will cause delays in providing the output. So, machine learning significantly depends on the data and its quality.

2. Time and Resources

The data that machines process remains huge in quantity and differs greatly. Machines require time so that their algorithm can adjust to the environment and learn it. Trials runs are held to check the accuracy and reliability of the machine. It requires massive and expensive resources and high-quality expertise to set up that quality of infrastructure. Trials runs are costly as they would cost in terms of time and expenses.

3. Results Interpretations

One of the biggest advantages of Machine learning is that interpreted data that we get from the cannot be hundred percent accurate. It will have some degree of inaccuracy. For a high degree of accuracy, algorithms should be developed so that they give reliable results.

4. High Error Chances

The error committed during the initial stages is huge, and if not corrected at that time, it creates havoc. Biasness and wrongness have to be dealt with separately; they are not interconnected. Machine learning depends on two factors, **i.e., data and algorithm**. All the errors are dependent on the two variables. Any incorrectness in any variables would have huge repercussions on the output.

5. Social Changes

Machine learning is bringing numerous social changes in society. The role of machine learning-based technology in society has **i**ncreased multifold. It is influencing the thought process of society and creating unwanted problems in

society. Character assassination and sensitive details are disturbing the social fabric of society.

6. Elimination of Human Interface

Automation, Artificial Intelligence, and Machine Learning have eliminated human interface from some work. It has eliminated employment opportunities. Now, all those works are conducted with the help of artificial intelligence and machine learning.

7. Changing Nature of Jobs

With the advancement of machine learning, the nature of the job is changing. Now, all the work are done by machine, and it is eating up the jobs for human which were done earlier by them. It is difficult for those without technical education to adjust to these changes.

8. Highly Expensive

This software is highly expensive, and not everybody can own it. Government agencies, big private firms, and enterprises mostly own it. It needs to be made accessible to everybody for wide use.

9. Privacy Concern

As we know that one of the pillars of machine learning is data. The collection of data has raised the fundamental question of privacy. The way data is collected and used for commercial purposes has always been a **contentious issue**. In India, the Supreme court of India has declared privacy a fundamental right of Indians. Without the user's permission, data cannot be collected, used, or stored. However, many cases have come up that big firms collect the data without the user's knowledge and using it for their commercial gains.

10. Research and Innovations

Machine learning is evolving concept. This area has not seen any major developments yet that fully revolutionized any economic sector. The area requires continuous research and innovation.

Applications of Machine Learning

- Image Recognition
- Speech Recognition
- Recommender Systems

- Fraud Detection
- Self Driving Cars
- Medical Diagnosis

2.Deep Learning

Deep learning is a multilayered, algorithmic technique in machine learning. The human brain's network of neurons is the inspiration for deep learning. Deep learning architecture plays an important role in perfecting the information that an AI system may process. The word 'deep' refers to the number of layers through which data transformation happens during the process.

Advantages of Deep Learning:

Deep learning has several advantages over traditional machine learning methods, some of the main ones include:

1. **Automatic feature learning:** Deep learning algorithms can automatically learn features from the data, which means that they don't require the features to be hand-engineered. This is particularly useful for tasks where the features are difficult to define, such as image recognition.
2. **Handling large and complex data:** Deep learning algorithms can handle large and complex datasets that would be difficult for traditional machine learning algorithms to process. This makes it a useful tool for extracting insights from big data.
3. **Improved performance:** Deep learning algorithms have been shown to achieve state-of-the-art performance on a wide range of problems, including image and speech recognition, natural language processing, and computer vision.
4. **Handling non-linear relationships:** Deep learning can uncover non-linear relationships in data that would be difficult to detect through traditional methods.
5. **Handling structured and unstructured data:** Deep learning algorithms can handle both structured and unstructured data such as images, text, and audio.
6. **Predictive modeling:** Deep learning can be used to make predictions about future events or trends, which can help organizations plan for the future and make strategic decisions.
7. **Handling missing data:** Deep learning algorithms can handle missing data and still make predictions, which is useful in real-world applications where data is often incomplete.

8. **Scalability:** Deep learning models can be easily scaled to handle an increasing amount of data and can be deployed on cloud platforms and edge devices.

Disadvantages of Deep Learning:

While deep learning has many advantages, there are also some disadvantages to consider:

1. **High computational cost:** Training deep learning models requires significant computational resources, including powerful GPUs and large amounts of memory. This can be costly and time-consuming.
2. **Overfitting:** Overfitting occurs when a model is trained too well on the training data and performs poorly on new, unseen data. This is a common problem in deep learning, especially with large neural networks, and can be caused by a lack of data, a complex model, or a lack of regularization.
3. **Lack of interpretability:** Deep learning models, especially those with many layers, can be complex and difficult to interpret. This can make it difficult to understand how the model is making predictions and to identify any errors or biases in the model.
4. **Dependence on data quality:** Deep learning algorithms rely on the quality of the data they are trained on. If the data is noisy, incomplete, or biased, the model's performance will be negatively affected.
5. **Data privacy and security concerns:** As deep learning models often rely on large amounts of data, there are concerns about data privacy and security. Misuse of data by malicious actors can lead to serious consequences like identity theft, financial loss and invasion of privacy.
6. **Lack of domain expertise:** Deep learning requires a good understanding of the domain and the problem you are trying to solve. If the domain expertise is lacking, it can be difficult to formulate the problem and select the appropriate algorithm.

Application of Deep Learning:

Deep learning has a wide range of applications across multiple industries and fields. Some of the most common applications include:

1. **Computer vision:** Deep learning is used in image and video recognition, object detection, semantic segmentation, and other

computer vision tasks. Applications include self-driving cars, security cameras, and image recognition for mobile devices.

2. **Natural language processing:** Deep learning is used in natural language understanding, machine translation, sentiment analysis, and other natural language processing tasks. Applications include chatbots, virtual assistants, and language-based search engines.

3. **Speech recognition:** Deep learning is used in speech recognition, voice identification, and voice synthesis. Applications include voice-controlled assistants, voice-enabled devices and voice-controlled robots.

4. **Predictive analytics:** Deep learning is used to analyze historical data and make predictions about future events. Applications include fraud detection, customer churn prediction, and demand forecasting.

5. **Recommender systems:** Deep learning is used to analyze patterns in data to recommend items to users. Applications include movie and music recommendations.

3. What is Artificial Intelligence:

Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing, speech recognition and machine vision.

Advantages of Artificial Intelligence (AI)

o **Reduction in Human Error**

One of the biggest achievements of Artificial Intelligence is that it can reduce human error. Unlike humans, a computer machine can't make mistakes if programmed correctly, while humans make mistakes from time to time. Therefore, Artificial Intelligence uses some set of algorithms by gathering previously stored data, reducing the chances of error and increasing the accuracy and precision of any task. Hence, Artificial Intelligence helps to solve complex problems that require difficult calculations and can be done without any error.

- **Reduce the Risk (Zero Risk)**

It is also one of the biggest advantages of Artificial Intelligence. The technology of developing AI Robots can overcome many risky limitations of humans and do risky things for us such as *defusing a bomb, oil and coal mining and exploring the deepest part of the ocean,* etc. So, it helps in any worst situation, either human or natural disasters too. AI Robots can be used in such situations where intervention can be hazardous.

- **24/7 Support**

Unlike humans, a computer does not require breaks and refreshers. A normal human can continue work till 8-9 hours, including breaks and refreshers, while a computer machine can work 24x7 without any breaks and don't even get bored, unlike humans. Chatbots and helpline centres can be seen as the best example of 24/7 support of various websites continuously engaged in receiving customers queries and automatically resolved by Artificial Intelligence.

- **Perform Repetitive Jobs**

We perform so many repetitive works in our day-to-day life, such as automatic replies to emails, sending birthday and anniversary quotes and verifying documents, etc. Therefore, Artificial Intelligence (AI) helps to automate the business by performing these repetitive jobs.

Disadvantages of Artificial Intelligence (AI)

o **High production cost**

We are living in a technological world where we have to manipulate ourselves according to society. Similarly, a computer machine also requires time to time software and hardware updates to meet the latest requirements. Hence, AI also need repairing and maintenance, which need plenty of costs.

o **Risk of Unemployment**

A robot is one of the implementations of Artificial intelligence, and it is replacing jobs and leading to serve unemployment (In some cases). Hence, according to some people, there is always a risk of unemployment because of robots and chatbots instead of humans. For example, in some more technology-oriented countries such as Japan, robots are widely used in manufacturing industries to replace human resources. However, this is not always the truth because as it replaces humans to enhance efficiency, it is also making more jobs opportunities for humans.

o **Increasing human's laziness**

The new inventions of Artificial Intelligence are making humans lazier towards their work, resulting in humans being completely dependent on machines and robots. If this continues for more upcoming years, then our next generations will become entirely dependent on a machine, resulting in further unemployment and health issues.

o **Emotionless**

We have always learned since childhood that computers or machines don't have emotions. Humans work like a team, and team management is a key factor for completing a target. However, there is no doubt that machines are much better when working efficiently, but it is also true that they never replace the human's connection that makes the team.

o **Lack of creativity**

The biggest disadvantage of Artificial Intelligence is its lack of creativity. Artificial Intelligence is a technology that is completely based on pre-loaded data. However, Artificial Intelligence can learn over time with this pre-fed data and past experiences, but it cannot be creative like humans.

Application of AI

Artificial Intelligence has various applications in today's society. It is becoming essential for today's time because it can solve complex problems with an efficient way in multiple industries, such as Healthcare, entertainment, finance, education, etc. AI is making our daily life more comfortable and fast.

Following are some sectors which have the application of Artificial Intelligence:
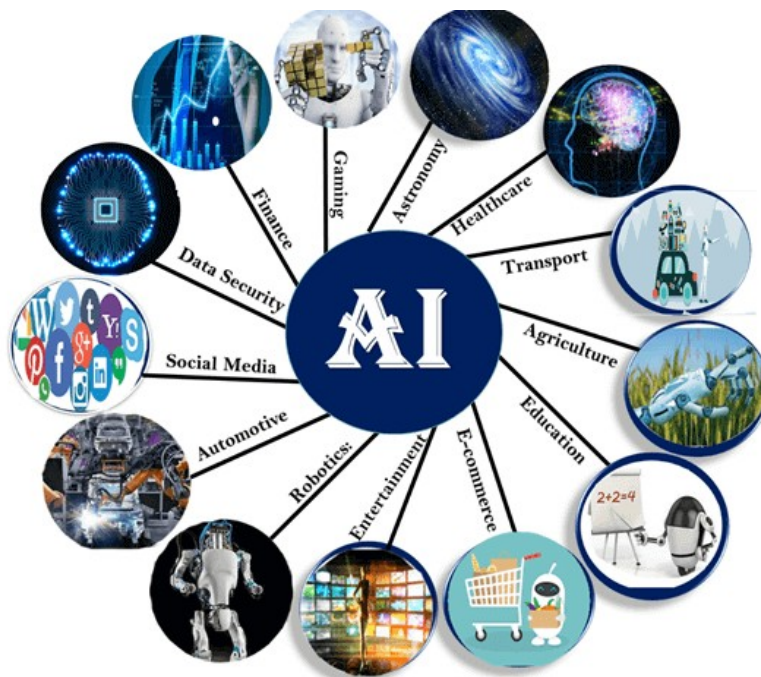
1. AI in Astronomy

o Artificial Intelligence can be very useful to solve complex universe problems. AI technology can be helpful for understanding the universe such as how it works, origin, etc.

2. AI in Healthcare

o In the last, five to ten years, AI becoming more advantageous for the healthcare industry and going to have a significant impact on this industry.

o   Healthcare Industries are applying AI to make a better and faster diagnosis than humans. AI can help doctors with diagnoses and can inform when patients are worsening so that medical help can reach to the patient before hospitalization.

3. AI in Gaming

o   AI can be used for gaming purpose. The AI machines can play strategic games like chess, where the machine needs to think of a large number of possible places.

4. AI in Finance

o   AI and finance industries are the best matches for each other. The finance industry is implementing automation, chatbot, adaptive intelligence, algorithm trading, and machine learning into financial processes.

5. AI in Data Security

o   The security of data is crucial for every company and cyber-attacks are growing very rapidly in the digital world. AI can be used to make your data more safe and secure. Some examples such as AEG bot, AI2 Platform,are used to determine software bug and cyber-attacks in a better way.

6. AI in Social Media

o   Social Media sites such as Facebook, Twitter, and Snapchat contain billions of user profiles, which need to be stored and managed in a very efficient way. AI can organize and manage massive amounts of data. AI can analyze lots of data to identify the latest trends, hashtag, and requirement of different users.

7. AI in Travel & Transport

o   AI is becoming highly demanding for travel industries. AI is capable of doing various travel related works such as from making travel arrangement to suggesting the hotels, flights, and best routes to the customers. Travel industries are using AI-powered chatbots which can make human-like interaction with customers for better and fast response.

Types of Machine Learning

Based on the methods and way of learning, machine learning is divided into mainly four types, which are:

1. Supervised Machine Learning
2. Unsupervised Machine Learning
3. Semi-Supervised Machine Learning
4. Reinforcement Learning
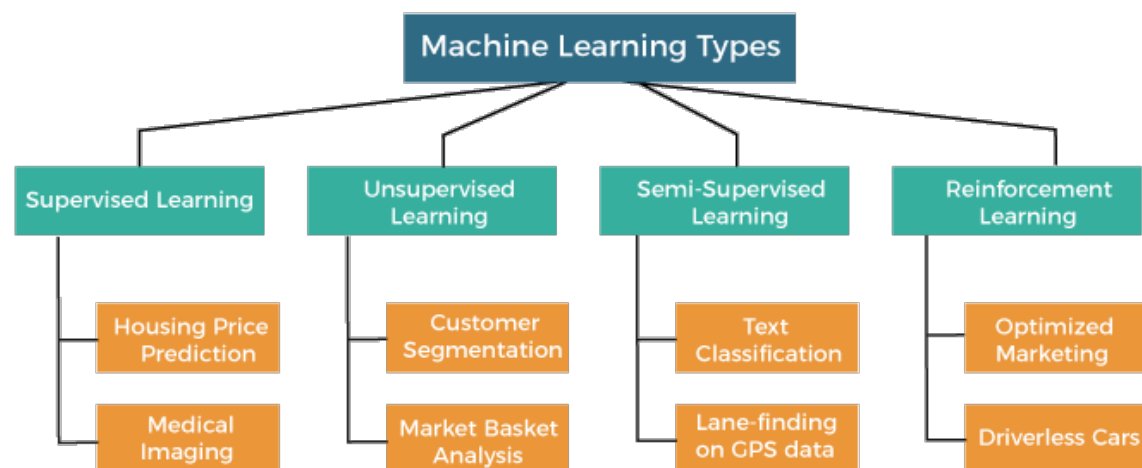
1. Supervised Machine Learning

As its name suggests, Supervised machine learning is based on supervision. It means in the supervised learning technique, we train the machines using the "labelled" dataset, and based on the training, the machine predicts the output. Here, the labelled data specifies that some of the inputs are already mapped to the output. More preciously, we can say; first, we train the machine with the input and corresponding output, and then we ask the machine to predict the output using the test dataset.

2. Unsupervised Machine Learning

Unsupervised learning is different from the Supervised learning technique; as its name suggests, there is no need for supervision. It means, in unsupervised machine learning, the machine is trained using the unlabeled dataset, and the machine predicts the output without any supervision.

In unsupervised learning, the models are trained with the data that is neither classified nor labelled, and the model acts on that data without any supervision.

3. Semi-Supervised Learning

Semi-Supervised learning is a type of Machine Learning algorithm that lies between Supervised and Unsupervised machine learning. It represents the intermediate ground between Supervised (With Labelled training data) and Unsupervised learning (with no labelled training data) algorithms and uses the combination of labelled and unlabeled datasets during the training period.

Although Semi-supervised learning is the middle ground between supervised and unsupervised learning and operates on the data that consists of a few labels, it mostly consists of unlabeled data. As labels are costly, but for corporate purposes, they may have few labels. It is completely different from supervised and unsupervised learning as they are based on the presence & absence of labels.

4. Reinforcement Learning

Reinforcement learning works on a feedback-based process, in which an AI agent (A software component) automatically explore its surrounding by hitting & trail, taking action, learning from experiences, and improving its performance. Agent gets rewarded for each good action and get punished for each bad action; hence the goal of reinforcement learning agent is to maximize the rewards.

In reinforcement learning, there is no labelled data like supervised learning, and agents learn from their experiences only.

5. Regression Model:

Regression analysis is a simple and statistical method to understand and quantify the relationship between two variables or more. It helps a business estimate one dependent variable based on the values of one or more independent variables.

To be precise, regression analysis helps individuals and businesses determine how changes in one variable are associated with changes in another. It's like finding a mathematical formula that best fits the data and allows to make predictions or understand the impact of different factors on an outcome

Regression Analysis Formula

1. Simple Linear regression formula: Simple linear regression is used when a single independent variable predicts a dependent variable. The linear regression formula is represented as $Y = a + bX$, where

Y is the dependent variable.

X is the independent variable.

a is the intercept (the value of Y when $X = 0$).

b is the slope (the change in Y for a one-unit change in X).

2. Multiple regression formula: Multiple regression extends linear regression by considering multiple independent variables to predict the dependent variable. The relationship is represented as $Y = a + b_1X_1 + b_2X_2 + ... + b_nX_n$, where

Y is the dependent variable.

$X_1, X_2, ..., X_n$ are the independent variables.

a is the intercept.

$b_1, b_2, ..., b_n$ are the coefficients of the independent variables.

3. Nonlinear regression formula: It is used in cases where the relationship between the dependent and independent variables is nonlinear. The model can take various forms depending on the specific problem. It is generally represented as $Y = f(X, \theta)$, where

Y is the dependent variable.

X is the independent variable(s).

θ represents the parameters of the nonlinear function f.

Types of Regression Analysis

Simple Linear Regression

Purpose: Simple linear regression is used to model the relationship between two variables, where one is considered the independent variable (predictor) and the other is the dependent variable (outcome).

Business Application: It's frequently used to identify how a change in one variable will affect another. For example, predicting sales based on advertising expenditure or estimating employee productivity based on hours worked.

Multiple Linear Regression

Purpose: Multiple linear regression extends simple linear regression to model relationships between multiple independent variables and a single dependent variable.

Business Application: Businesses use it to understand how multiple factors influence outcomes. For instance, predicting home prices based on features like square footage, number of bedrooms, and neighborhood.

Logistic Regression

Purpose: Logistic regression is used when the dependent variable is binary (two possible outcomes). It models the probability of a particular outcome occurring.

Business Application: In business, logistic regression is employed for tasks like predicting customer churn (yes/no), whether a customer will purchase a product (yes/no), or whether a loan applicant will default on a loan (yes/no).

Polynomial Regression

Purpose: Polynomial regression is used when the relationship between the independent and dependent variables follows a polynomial curve and is not linear.

Business Application: It can be used to model more complex relationships in data, such as predicting the growth of a plant-based on time and other environmental factors.

6.concept of classification clustering and reinforcement learning

Classification

The aim of the classification is to split the data into two or more predefined groups. A common example is spam email filtering where emails are split into either spam or not spam.
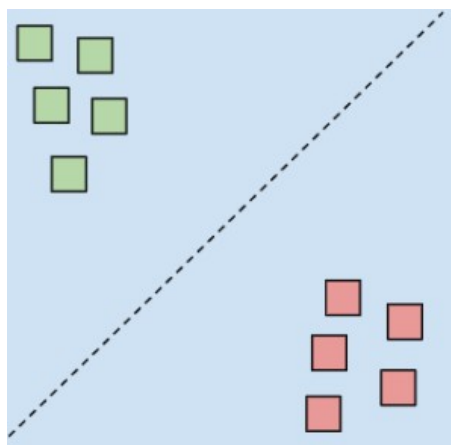

Clustering:

Clustering or cluster analysis is a machine learning technique, which groups the unlabelled dataset. It can be defined as *"A way of grouping the data points into different clusters, consisting of similar data points. The objects with the possible similarities remain in a group that has less or no similarities with another group."*

Below diagram explains the working of the clustering algorithm. We can see the different fruits are divided into several groups with similar properties.
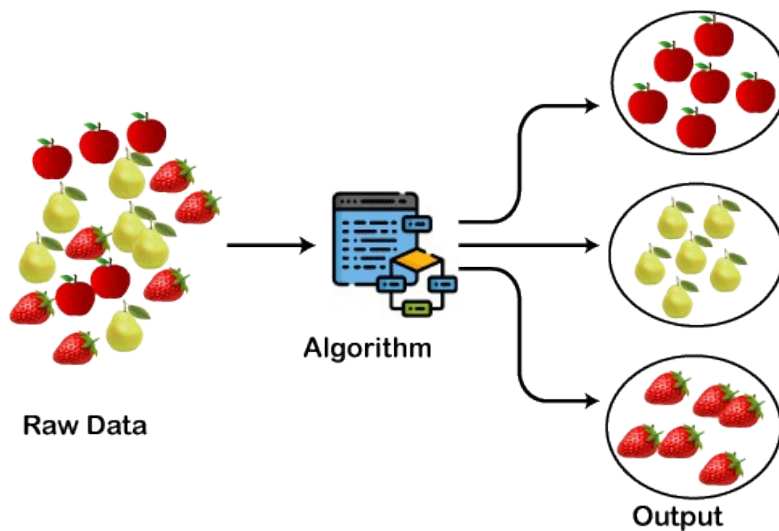
Reinforcement learning:

Reinforcement learning is an area of Machine Learning. It is about taking suitable action to maximize reward in a particular situation. It is employed by various software and machines to find the best possible behavior or path it should take in a specific situation. Reinforcement learning differs from supervised learning in a way that in supervised learning the training data has the answer key with it so the model is trained with the correct answer itself whereas in reinforcement learning, there is no answer but the reinforcement agent decides what to do to perform the given task. In the absence of a training dataset, it is bound to learn from its experience.
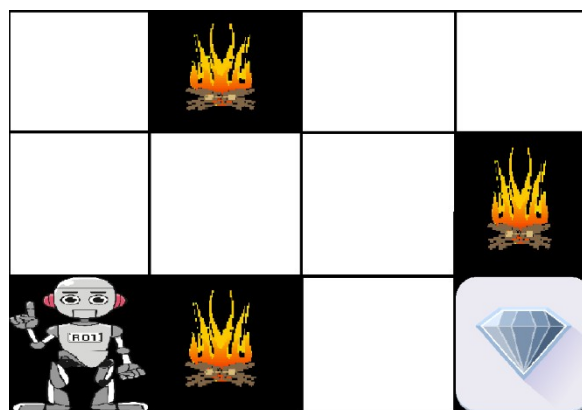
The above image shows the robot, diamond, and fire. The goal of the robot is to get the reward that is the diamond and avoid the hurdles that are fired. The robot learns by trying all the possible paths and then choosing the path which gives him the reward with the least hurdles. Each right step will give the robot a reward and each wrong step will subtract the reward of the robot. The total reward will be calculated when it reaches the final reward that is the diamond.