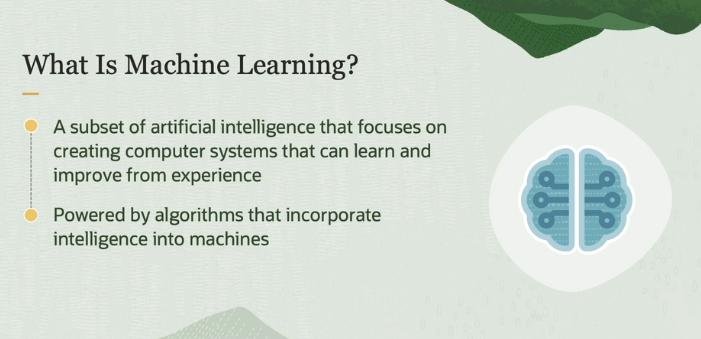
**Introduction to ML**

****

best resource : <https://www.ibm.com/topics/machine-learning>

**What is Machine Learning?**

* **Machine Learning** is a **subset of Artificial Intelligence (AI)** that deals with designing systems that can **learn from data** and **make predictions** or decisions **without explicit programming**. Instead of writing step-by-step instructions for a computer to follow, ML systems figure out patterns and rules by analyzing **examples (data)**.

**Key Concepts:**

1. **Learning and Predicting**:
   * ML systems are designed to **learn from past examples** (historical data) and use that learning to **predict outcomes** for new, unseen data.
   * **Example**: A spam filter learns from past emails that were labeled as "spam" or "not spam" and uses that knowledge to predict whether a new incoming email is spam.
2. **No Explicit Programming**:
   * Traditional programming involves providing **explicit instructions** on how to perform tasks. For example, in a rule-based system, you'd write all the possible rules for sorting spam emails.
   * In ML, however, the system doesn’t need specific rules; instead, it **learns from patterns** in the data itself. The system gets better at making decisions by analyzing examples, instead of being told what to do step-by-step.
3. **Powered by Algorithms**:
   * Machine learning relies on **algorithms**, which are mathematical models designed to analyze data, recognize patterns, and make predictions. These algorithms are the core tools that allow the machine to "learn."
   * **Example Algorithms**: Decision Trees, Neural Networks, Support Vector Machines (SVM), and K-Nearest Neighbors.
4. **Learning from Data**:
   * The **data** provided to an ML model acts as examples from which it learns. The more data you provide, the better the model becomes at identifying patterns and making accurate predictions.
   * **Example**: In a machine learning model trained to recognize handwriting, the model improves its ability to identify digits (like "2" or "7") the more examples of handwriting it is exposed to.

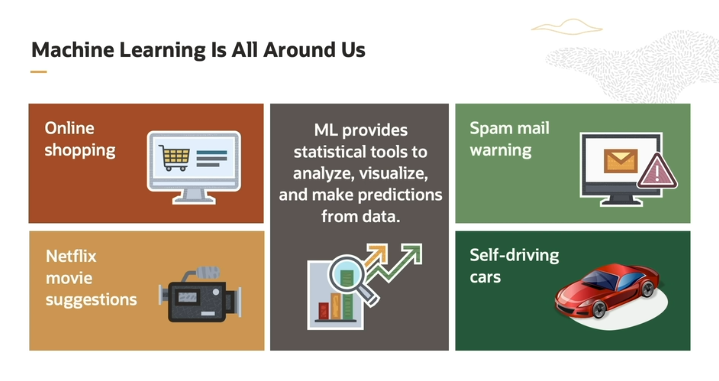
**Examples of Machine Learning in Action:**

1. **Image Recognition**:
   * ML models can be trained to recognize objects in images, such as identifying whether an image contains a cat or a dog. These models are widely used in facial recognition systems, medical imaging, and self-driving cars.
2. **Natural Language Processing (NLP)**:
   * Machine learning powers chatbots and virtual assistants like Siri and Alexa. These systems learn from large datasets of text and conversations, allowing them to understand and respond to spoken or written language.
3. **Recommendation Systems**:
   * Companies like Netflix and Amazon use ML models to **recommend** movies, products, or shows. The system learns from your past behavior (e.g., the shows you’ve watched) and predicts what you might want to watch next.
4. **Fraud Detection**:
   * In financial systems, ML models analyze transaction data to **detect fraudulent activities**. They learn patterns of normal transactions and flag unusual behaviors that could indicate fraud.

**Summary:**

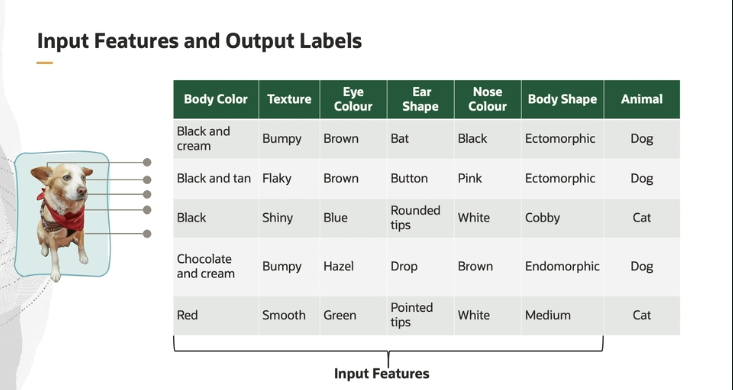
* **Machine Learning** is a field of AI where **computers learn from data** to make predictions or decisions, without requiring explicit instructions for each task.
* It is powered by **algorithms** that automatically identify patterns from examples provided as data.
* ML has numerous real-world applications, from **spam filtering** and **image recognition** to **fraud detection** and **product recommendations**.

In essence, machine learning allows computers to mimic human learning by **generalizing** from past experiences (data) and applying that knowledge to new situations.



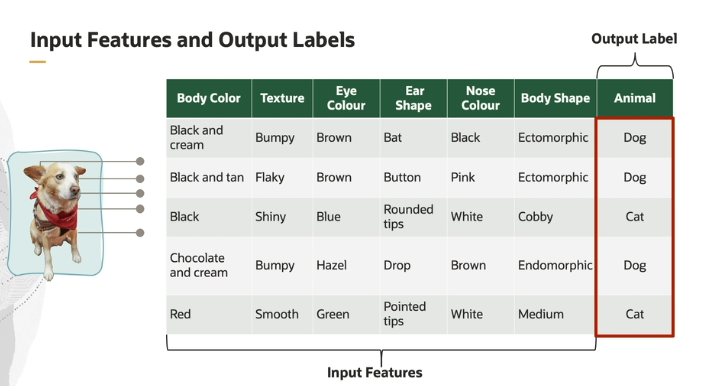
Machine learning is used by all of us in our day-to-day life. When we shop online, we get product recommendations based on our preferences and our shopping history. This is powered by machine learning. We are notified about movies recommendations on Netflix based on our viewing history and choices of other similar viewers. This too is driven by machine learning.

While browsing emails, we are warned of a spam mail because machine learning classifies whether the mail is spam or not based on its content. In the increasingly popular self-driving cars, machine learning is responsible for taking the car to its destination. Let us see how machine learning works.



Let us say we have a computer, and we need to teach the computer to differentiate between a cat and a dog. We do this by describing features of a cat or a dog. Dogs and cats have distinguishing features. For example, the body color, texture, eye color are some of the defining features which can be used to differentiate a cat from a dog. These are collectively called as input data.

We also provide a corresponding output, which is called as a label, which can be a dog or a cat in this case. By describing a specific set of features, we can see that it is a cat or a dog. Machine learning model is first trained with the data set. Training data set consists of a set of features and output labels and is given as an input to the machine learning model. During the process of training, machine learning model learns the relation between input features and corresponding output labels from the provided data.



So its mean for training the ML model we have to provide it 2 things:

**1- Input feature** -> basically the data about the object or data of any organization or data for any product for which you creating this model.

In **Machine Learning (ML)**, **input features** refer to the individual measurable properties or characteristics of the data that are used as input for a model to make predictions or perform tasks.

**Key Points about Input Features:**

1. **Definition**:
   * Input features (often called **attributes**, **variables**, or **predictors**) are the pieces of data or information that the machine learning model uses to learn from and make predictions.
   * In a dataset, each feature represents one aspect of the data.
2. **Types**:
   * **Numerical features**: Continuous or discrete values (e.g., age, temperature, number of products sold).
   * **Categorical features**: Variables that represent categories or labels (e.g., gender, type of product, color).
   * **Binary features**: Special case of categorical variables with only two possible values (e.g., yes/no, true/false).
3. **Examples**:
   * In a **housing price prediction model**, input features might include:
     + Square footage of the house
     + Number of bedrooms
     + Neighborhood
     + Year built
   * In a **fraud detection model**, input features might include:
     + Transaction amount
     + Time of transaction
     + Customer's previous transaction history
     + Location of the transaction
4. **Purpose**:
   * The input features represent the **factors** that influence the **target outcome** (the prediction the model is trying to make).
   * The ML model learns the relationship between input features and the target variable during training.
5. **Importance of Feature Selection**:
   * Not all features are equally useful. **Irrelevant or redundant features** can reduce model performance.
   * **Feature selection** is the process of identifying the most important or relevant features that contribute to the accuracy of the model.

In summary, input features are the essential pieces of information that the model uses to **train**, **learn patterns**, and ultimately make **predictions**.

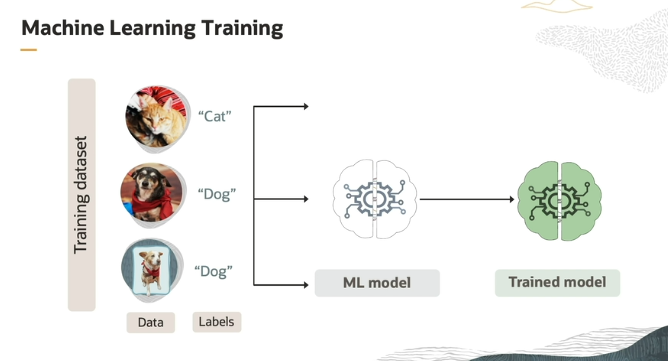
**2- Output label:** In **Machine Learning (ML) training**, the **output label** (often referred to as the **target** or **dependent variable**) is the value or category that the model is being trained to predict or classify based on the input features.

**Key Points about Output Labels:**

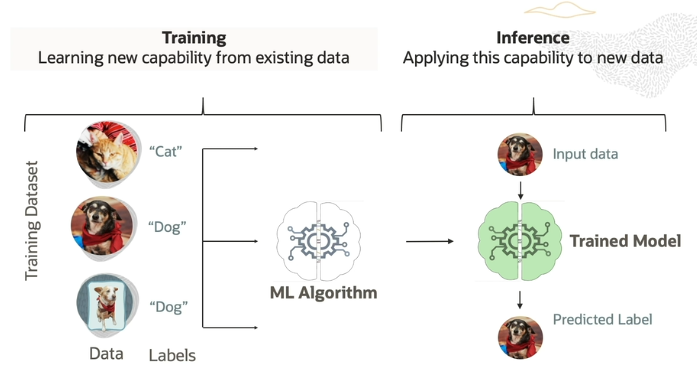
1. **Definition**:
   * The output label is the **correct answer** or **true result** that corresponds to a set of input features in the dataset.
   * It is the value the model tries to learn to predict during the training process.
2. **Types**:
   * **Numerical labels**: Used in **regression** tasks where the output is a continuous value (e.g., predicting house prices, temperature).
   * **Categorical labels**: Used in **classification** tasks where the output is a category or class (e.g., predicting if an email is spam or not, identifying whether an image is of a dog or a cat).
3. **Examples**:
   * In a **house price prediction** model:
     + **Input features**: square footage, number of rooms, location, etc.
     + **Output label**: the actual price of the house.
   * In a **spam email classifier**:
     + **Input features**: words used in the email, sender information, time sent, etc.
     + **Output label**: whether the email is **spam** or **not spam**.
   * In an **image classification** task:
     + **Input features**: pixel values of the image.
     + **Output label**: the category of the image (e.g., **cat**, **dog**, **car**, etc.).
4. **Purpose in Training**:
   * During training, the model is given input features and corresponding output labels. The model learns the relationship between the inputs and the correct output.
   * The model’s performance is evaluated by comparing its predictions to the actual output labels, and adjustments are made to minimize the difference (loss or error).
5. **Training Process**:
   * The **input features** are passed to the model, and the model generates a prediction.
   * This prediction is compared to the **output label** (the correct answer).
   * Based on this comparison, the model adjusts its internal parameters (using techniques like backpropagation) to improve its future predictions.

**Summary:**

The **output label** is what the ML model aims to predict. It is provided during the training phase to guide the learning process, allowing the model to adjust

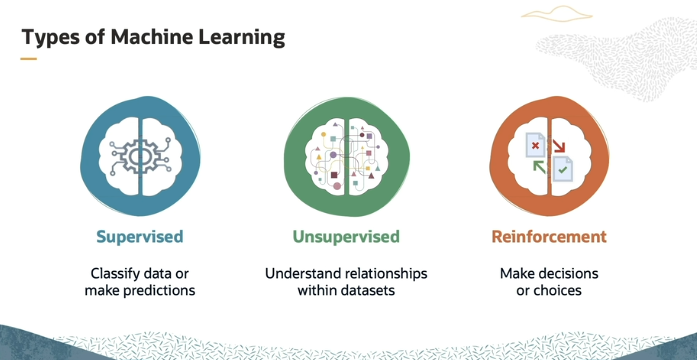


Now our model is trained !



Once the model learns from the data, we have a trained model. Once the model is trained, it can be used for inference. Inference is a process of getting a prediction by giving a data point. In this example, we input features of a cat or a dog, and the trained model predicts the output that is a cat or a dog label.

Now you can see in the above e.g that we have provided an input data to our **Trained Model** and it predicts an output.

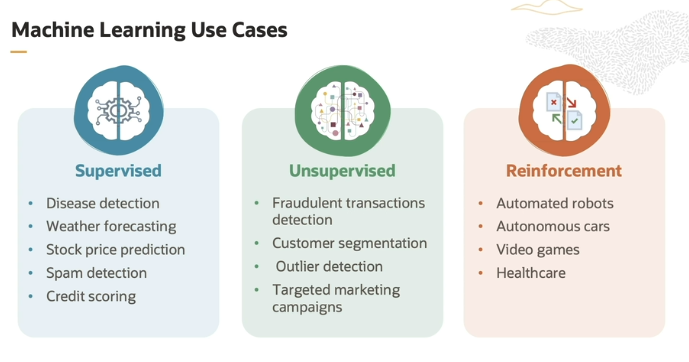
  
**Supervise Learning :** InSupervise learning, we provide a labeled Data as output to model and input features of Data. Like let say we are creating a model for detecting a bottle and jug. So first we provide input feature of Bottle and jug to model , and after that we will provide it output labels as well like for bottle we will write **bottle** against the input feature of bottle So that it can be able to detect whenever the input data is matched with input features so it will detect it as **Bottle** same goes for jug, Model learns the relation between features and labels . Now **Supervised ML** perform two task **classification(**its mean kay wo detect/identify krta hai on the basis of given input.**)** and **Predictions/Inference (**and after doing classification usko future perspective ka A/C predict krna or infer krna and then final.

**Unsupervised Learning:**

Unsupervised learning is generally used to understand relationships within a data set. Labels are not used or are not available**.** Its is used when you dont have exact label data. Like if you are creating a model for ecommerce store ab usme apko product recommendation ka liya model create krna hai toh Zahir hai usme toh koi label data nhi hai so that’s why we wil use unsupervised ML in these situations.

**Reinforcement ML:**

Reinforcement learning uses algorithms that learn from outcomes to make decisions or choices. Let us talk about some real examples of machine learning. Isma simply model ko ek certain amount of necessary dataset pa train krdetay hain, then model ma jo further improvements krni hain , it will done by performing action and learning from the feedback of action.



Some of the popular applications of supervised machine learning are disease detection, weather forecasting, stock price prediction, spam detection, and credit scoring. For example, in disease detection, the patient data is input to a machine learning model, and machine learning model predicts if a patient is suffering from a disease or not.

For unsupervised machine learning, some of the most common real-time applications are to detect fraudulent transactions, customer segmentation, outlier detection, and targeted marketing campaigns. So, for example, given the transaction data, we can look for patterns that lead to fraudulent transactions.

Most popular among reinforcement learning applications are automated robots, autonomous driving cars, and playing games

**Feature engineering:**

Feature engineering is the process of selecting, manipulating and transforming raw data into features that can be used in [supervised learning](https://builtin.com/machine-learning/supervised-learning). It’s also necessary to design and train new [machine learning](https://builtin.com/machine-learning/machine-learning-basics) features so it can tackle new tasks. A “feature” is any measurable input that can be used in a [predictive model](https://builtin.com/machine-learning/predictive-behavior-modeling). It could be the color of an object or the sound of someone’s voice. Feature engineering is the act of converting raw observations into desired features using statistical or machine learning approaches.

In simple terms , as we know that when user pass the data to our model so it is a raw data BUT our model need the features to identify the data and make predictions or classification on that input . So the process of converting

**Resources:** [**https://www.ibm.com/topics/feature-engineering**](https://www.ibm.com/topics/feature-engineering)