**1.Explain the term machine learning, and how does it work? Explain two machine learning applications in the business world. What are some of the ethical concerns that machine learning applications could raise?**

Ans : Machine Learning is an Application of Artificial Intelligence (AI) it gives devices the ability to learn from their experiences and improve their self without doing any coding.

Diagram

Description automatically generated

There are Seven Steps of Machine Learning

* Gathering Data
* Preparing that data
* Choosing a model
* Training
* Evaluation
* Hyperparameter Tuning
* Prediction

Business Use cases :

1. Facial recognition/Image recognition

The most common application of machine learning is Facial Recognition, and the simplest example of this application is the iPhone X. There are a lot of use-cases of facial recognition, mostly for security purposes like identifying criminals, searching for missing individuals, aid forensic investigations, etc. Intelligent marketing, diagnose diseases, track attendance in schools, are some other uses.

1. Financial Services

Machine learning has many use cases in Financial Services. Machine Learning algorithms prove to be excellent at detecting frauds by monitoring activities of each user and assess that if an attempted activity is typical of that user or not.

Financial monitoring to detect money laundering activities is also a critical security use case of machine learning.

Machine Learning also helps in making better trading decisions with the help of algorithms that can analyse thousands of data sources simultaneously. Credit scoring and underwriting are some of the other applications.

The most common application in our day to day activities is the virtual personal assistants like Siri and Alexa.

The ethical issues that were identified :

* Cost to innovation.
* Harm to physical integrity.
* Lack of access to public services.
* Lack of trust.
* “Awakening” of AI.
* Security problems.
* Lack of quality data.
* Disappearance of jobs.

**2. Describe the process of human learning:**

**i. Under the supervision of experts**

**ii. With the assistance of experts in an indirect manner**

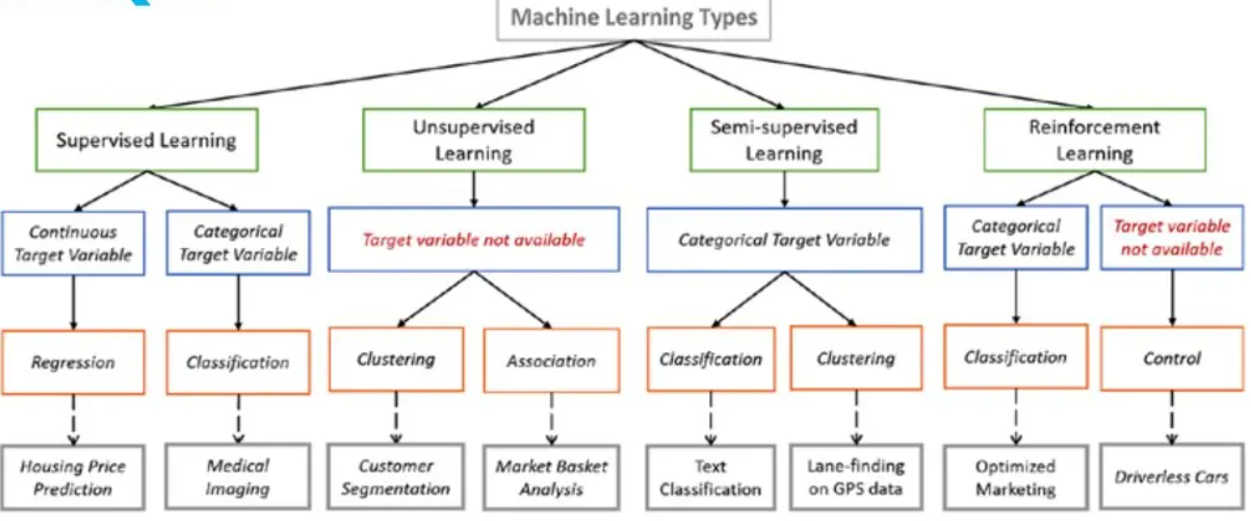
**iii. Self-education**

Ans : Learning is the process of acquiring new understanding, knowledge, behaviors, skills, values, attitudes, and preferences.

Learning is a key process in human behaviour. All living is learning. If we compare the simple, crude ways in which a child feels and behaves, with the complex modes of adult behaviour, his skills, habits, thought, sentiments and the like- we will know what difference learning has made to the individual.The individual is constantly interacting with and influenced by the environment. This experience makes him to change or modify his behaviour in order to deal effectively with it. Therefore, learning is a change in behaviour, influenced by previous behaviour. As stated above the skills, knowledge, habits, attitudes, interests and other personality characteristics are all the result of learning.Learning is defined as “any relatively permanent change in behaviour that occurs as a result of practice and experience”. This definition has three important elements.

**3. Provide a few examples of various types of machine learning.**

Ans : These are three types of machine learning: supervised learning, unsupervised learning, and reinforcement learning.



**4. Examine the various forms of machine learning.**

Ans :

**Diagram

Description automatically generated**

**5. Can you explain what a well-posed learning problem is? Explain the main characteristics that must be present to identify a learning problem properly.**

Ans : **Well Posed Learning Problem –** A computer program is said to learn from experience E in context to some task T and some performance measure P, if its performance on T, as was measured by P, upgrades with experience E.

Any problem can be segregated as well-posed learning problem if it has three traits –

* Task
* Performance Measure
* Experience

**Certain examples that efficiently defines the**well-posed**learning problem are –**

**1. To better filter emails as spam or not**

* Task – Classifying emails as spam or not
* Performance Measure – The fraction of emails accurately classified as spam or not spam
* Experience – Observing you label emails as spam or not spam

**2. A checkers learning problem**

* Task – Playing checkers game
* Performance Measure – percent of games won against opposer
* Experience**–** playing implementation games against itself

**3. Handwriting Recognition Problem**

* Task – Acknowledging handwritten words within portrayal
* Performance Measure – percent of words accurately classified
* Experience – a directory of handwritten words with given classifications

**4. A Robot Driving Problem**

* Task – driving on public four-lane highways using sight scanners
* Performance Measure – average distance progressed before a fallacy
* Experience – order of images and steering instructions noted down while observing a human driver

**5. Fruit Prediction Problem**

* Task – forecasting different fruits for recognition
* Performance Measure – able to predict maximum variety of fruits
* Experience – training machine with the largest datasets of fruits images

**6. Face Recognition Problem**

* Task – predicting different types of faces
* Performance Measure – able to predict maximum types of faces
* Experience – training machine with maximum amount of datasets of different face images

**6. Is machine learning capable of solving all problems? Give a detailed explanation of your answer.**

Ans :

**ML Can't Solve Everything.** **Here Are 5 Challenges That It Still Faces**

* Reasoning Power. One area where ML has not mastered successfully is reasoning power, a distinctly human trait.
* Contextual Limitation.
* Scalability.
* Regulatory Restriction For Data In ML.
* Internal Working Of Deep Learning.

**7. What are the various methods and technologies for solving machine learning problems? Any two of them should be defined in detail.**

Ans : The ten methods described offer an overview — and a foundation you can build on as you hone your machine learning knowledge and skill:

1. **Regression**
2. **Classification**
3. **Clustering**
4. **Dimensionality Reduction**
5. **Ensemble Methods**
6. **Neural Nets and Deep Learning**
7. **Transfer Learning**
8. **Reinforcement Learning**
9. **Natural Language Processing**
10. **Word Embeddings**

Regression

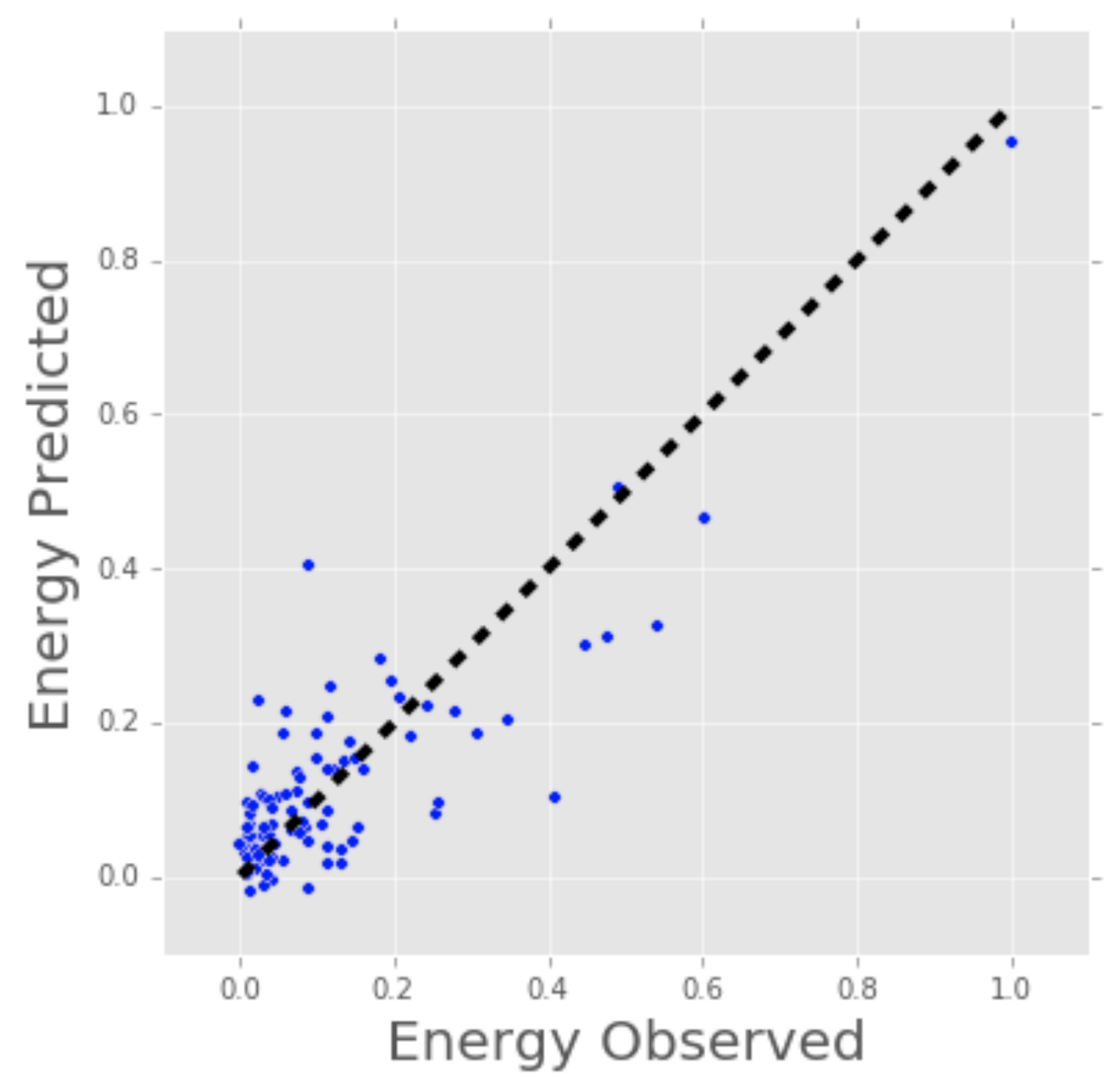
Regression methods fall within the category of supervised ML. They help to predict or explain a particular numerical value based on a set of prior data, for example predicting the price of a property based on previous pricing data for similar properties.

The simplest method is linear regression where we use the mathematical equation of the line (**y = m \* x + b**) to model a data set. We train a linear regression model with many data pairs **(x, y)** by calculating the position and slope of a line that minimizes the total distance between all of the data points and the line. In other words, we calculate the slope (**m**) and the y-intercept (**b**) for a line that best approximates the observations in the data.

Let’s consider a more a concrete example of linear regression. I once used a linear regression to predict the energy consumption (in kWh) of certain buildings by gathering together the age of the building, number of stories, square feet and the number of plugged wall equipment. Since there were more than one input (age, square feet, etc…), I used a multi-variable linear regression. The principle was the same as a simple one-to-one linear regression, but in this case the “line” I created occurred in multi-dimensional space based on the number of variables.

The plot below shows how well the linear regression model fit the actual energy consumption of building. Now imagine that you have access to the characteristics of a building (age, square feet, etc…) but you don’t know the energy consumption. In this case, we can use the fitted line to approximate the energy consumption of the particular building.

Note that you can also use linear regression to estimate the weight of each factor that contributes to the final prediction of consumed energy. For example, once you have a formula, you can determine whether age, size, or height is most important.



Linear Regression Model Estimates of Building’s Energy Consumption (kWh).

Regression techniques run the gamut from simple (like linear regression) to complex (like regularized linear regression, polynomial regression, decision trees and random forest regressions, neural nets, among others). But don’t get bogged down: start by studying simple linear regression, master the techniques, and move on from there.

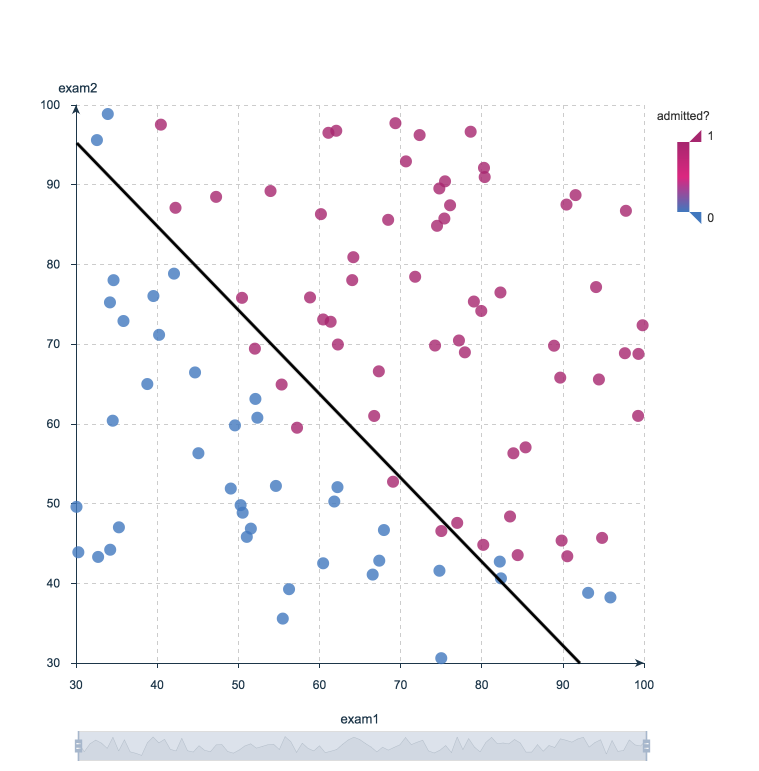
Classification

Another class of supervised ML, classification methods predict or explain a class value. For example, they can help predict whether or not an online customer will buy a product. The output can be yes or no: buyer or not buyer. But classification methods aren’t limited to two classes. For example, a classification method could help to assess whether a given image contains a car or a truck. In this case, the output will be 3 different values: 1) the image contains a car, 2) the image contains a truck, or 3) the image contains neither a car nor a truck.

The simplest classification algorithm is logistic regression — which makes it sounds like a regression method, but it’s not. Logistic regression estimates the probability of an occurrence of an event based on one or more inputs.

For instance, a logistic regression can take as inputs two exam scores for a student in order to estimate the probability that the student will get admitted to a particular college. Because the estimate is a probability, the output is a number between 0 and 1, where 1 represents complete certainty. For the student, if the estimated probability is greater than 0.5, then we predict that he or she will be admitted. If the estimated probabiliy is less than 0.5, we predict the he or she will be refused.

The chart below plots the scores of previous students along with whether they were admitted. Logistic regression allows us to draw a line that represents the decision boundary.



Logistic Regression Decision Boundary: Admitted to College or Not?

Because logistic regression is the simplest classification model, it’s a good place to start for classification. As you progress, you can dive into non-linear classifiers such as decision trees, random forests, support vector machines, and neural nets, among others.

**8. Can you explain the various forms of supervised learning? Explain each one with an example application.**

Ans : **Different Types of Supervised Learning**

**1. Regression**

In regression, a single output value is produced using training data. This value is a probabilistic interpretation, which is ascertained after considering the strength of correlation among the input variables. For example, regression can help predict the price of a house based on its locality, size, etc.

In logistic regression, the output has discrete values based on a set of independent variables. This method can flounder when dealing with non-linear and multiple decision boundaries. Also, it is not flexible enough to capture complex relationships in datasets.

**2. Classification**

It involves grouping the data into classes.  If you are thinking of extending credit to a person, you can use classification to determine whether or not a person would be a loan defaulter. When the supervised learning algorithm labels input data into two distinct classes, it is called binary classification. Multiple classifications means categorizing data into more than two classes.

**3. Naive Bayesian Model**

The Bayesian model of classification is used for large finite datasets. It is a method of assigning class labels using a direct acyclic graph. The graph comprises one parent node and multiple children nodes. And each child node is assumed to be independent and separate from the parent.

**Decision Trees**

A decision tree is a flowchart-like model that contains conditional control statements, comprising decisions and their probable consequences. The output relates to the labelling of unforeseen data.

In the tree representation, the leaf nodes correspond to class labels, and the internal nodes represent the attributes. A decision tree can be used to solve problems with discrete attributes as well as boolean functions. Some of the notable decision tree algorithms are ID3 and CART.

**4. Random Forest Model**

The random forest model is an ensemble method. It operates by constructing a multitude of decision trees and outputs a classification of the individual trees. Suppose you want to predict which undergraduate students will perform well in GMAT – a test taken for admission into graduate management programs. A random forest model would accomplish the task, given the demographic and educational factors of a set of students who have previously taken the test.

**5. Neural Networks**

This algorithm is designed to cluster raw input, recognize patterns, or interpret sensory data. Despite their multiple advantages, neural networks require significant computational resources. It can get complicated to fit a neural network when there are thousands of observations. It is also called the ‘black-box’ algorithm as interpreting the logic behind their predictions can be challenging.

**6. Support Vector Machines**

Support Vector Machine (SVM) is a supervised learning algorithm developed in the year 1990. It draws from the statistical learning theory developed by Vap Nick.

SVM separates hyperplanes, which makes it a discriminative classifier. The output is produced in the form of an optimal hyperplane that categorizes new examples. SVMs are closely connected to the kernel framework and used in diverse fields. Some examples include bioinformatics, pattern recognition, and multimedia information retrieval.

**9. What is the difference between supervised and unsupervised learning? With a sample application in each region, explain the differences.**

Ans : The main difference between supervised and unsupervised learning: Labeled data

The main distinction between the two approaches is the use of labeled datasets. To put it simply, supervised learning uses labeled input and output data, while an unsupervised learning algorithm does not.

In supervised learning, the algorithm “learns” from the training dataset by iteratively making predictions on the data and adjusting for the correct answer. While supervised learning models tend to be more accurate than unsupervised learning models, they require upfront human intervention to label the data appropriately. For example, a supervised learning model can predict how long your commute will be based on the time of day, weather conditions and so on. But first, you’ll have to train it to know that rainy weather extends the driving time.

Unsupervised learning models, in contrast, work on their own to discover the inherent structure of unlabeled data. Note that they still require some human intervention for validating output variables. For example, an unsupervised learning model can identify that online shoppers often purchase groups of products at the same time. However, a data analyst would need to validate that it makes sense for a recommendation engine to group baby clothes with an order of diapers, applesauce and sippy cups.

Other key differences between supervised and unsupervised learning

* **Goals:**In supervised learning, the goal is to predict outcomes for new data. You know up front the type of results to expect. With an unsupervised learning algorithm, the goal is to get insights from large volumes of new data. The machine learning itself determines what is different or interesting from the dataset.
* **Applications**: Supervised learning models are ideal for spam detection, sentiment analysis, weather forecasting and pricing predictions, among other things. In contrast, unsupervised learning is a great fit for anomaly detection, recommendation engines, customer personas and medical imaging.
* **Complexity:**Supervised learning is a simple method for machine learning, typically calculated through the use of programs like R or Python.In unsupervised learning, you need powerful tools for working with large amounts of unclassified data. Unsupervised learning models are computationally complex because they need a large training set to produce intended outcomes.
* **Drawbacks**: Supervised learning models can be time-consuming to train, and the labels for input and output variables require expertise. Meanwhile, unsupervised learning methods can have wildly inaccurate results unless you have human intervention to validate the output variables.

**10. Describe the machine learning process in depth.**

**a. Make brief notes on any two of the following:**

**MATLAB is one of the most widely used programming languages.**

**ii. Deep learning applications in healthcare**

**iii. Study of the market basket**

**iv. Linear regression (simple)**

**11. Make a comparison between:-**

**1. Generalization and abstraction**

**2. Learning that is guided and unsupervised**

**3. Regression and classification**