Linear and logistic regression are traditional way of doing things which uses statistics as base to predict the dependent variable. Statistical Modelling find of relationships between variables in the form of mathematical equations. But they are drawbacks to these algorithms. Few of them are listed below

* Statistical Modelling has to satisfy the assumptions that are discussed previously. If they are not satisfied, models won’t be reliable and through random predictions
* As there are more assumptions, predictive power of the model will be low
* These algorithm face challenge when data is non-linear. Complex patterns are hard to decode.
* Data should be clean (Missing values and outliers should not be present)

To overcome all these limitations came concept of Machine Learning.

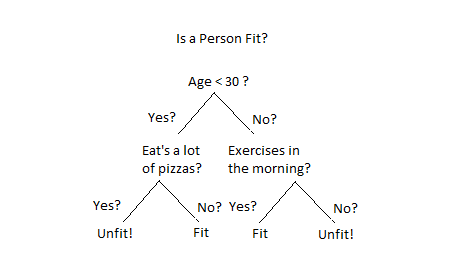
**Machine Learning is** *a subfield artificial intelligence which deals with building systems that can learn from data and try to achieve goals using different techniques. ML algorithms don’t try to find the mathematical equation instead search for pattern to the problem in hand with better accuracy possible. The objective is to make computers learn on its own without any intervention from the human.*

Let’s get into each algorithm and see how it works to solve the problem

**Decision Tree**

Decision is a type of supervised learning, in which the data is split in to similar groups based on the based on the most important variable to the least. It looks like a tree shape once all the variables split hence the name tree-based models.

The structure is a tree with decision nodes and leaf nodes. A decision node can have two or more branches and leaf node represents a decision. Decision trees can handle both categorical and numerical data.



How does a tree decide where to split?

The core of decision tree algorithm is the process of splitting the trees. The decision of making splits heavily affects a tree’s accuracy. So it uses different algorithms to split the node and it’s also different for classification and regression problems. Let’s look at each one of those separately

1. Classification Problems:
2. Gini Index