KNN

- The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.

- A **supervised machine learning** algorithm (as opposed to an unsupervised machine learning algorithm) is one that relies on labeled input data to learn a function that produces an appropriate output when given new unlabeled data.

- Supervised machine learning algorithms are used to solve classification or regression problems.

- A **classification problem** has a discrete value as its output.

- A **regression problem** has a real number (a number with a decimal point) as its output.

- We have an independent variable (or set of independent variables) and a dependent variable (the thing we are trying to guess given our independent variables).

- An **unsupervised machine learning** algorithm makes use of input data without any labels —in other words, no teacher (label) telling the child (computer) when it is right or when it has made a mistake so that it can self-correct.

- **KNN** is a **non-parametric, lazy** learning algorithm**.**

- The training phase is pretty fast

- KNN Algorithm is based on **feature similarity**

When do we use KNN Algorithm?

KNN can be used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry. To evaluate any algorithms, we generally look at 3 important aspects:

- Ease to interpret output
- Calculation time

- Predictive Power

- KNN is still the better choice for applications where predictions are not

requested frequently but where accuracy is important.

Pros and Cons of KNN

Pros:

- It is a very simple algorithm to understand and interpret.
- It is very useful for nonlinear data because there is no assumption about data in this algorithm.

- It is a versatile algorithm as we can use it for classification as well as regression.
- It has relatively high accuracy

Cons

- It is computationally a bit expensive algorithm because it stores all the training data.

- High memory storage required as compared to other supervised learning algorithms.

- Prediction is slow in the case of big N.

- It is very sensitive to the scale of data as well as irrelevant features.

Support Vector Classifier

- Support vector machine is another simple algorithm

- Support Vector Machine" (SVM) is a supervised Machine Learning algorithm that can be used for both classification or regression challenges.

- The support vector machine is highly preferred by many as it produces significant accuracy with less computation power.
- SVM   can be used for both regression and classification tasks.
- The objective of the support vector machine algorithm is to find a   hyperplane in N-dimensional space (N — the number of features) that distinctly classifies the data points.

- To separate the two classes of data points, there are many possible hyperplanes that could be chosen.

- Our objective is to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes.

- Maximizing the margin distance provides some reinforcement so that future data

points can be classified with more confidence.