# Weekly Report 2 - KNN

Ganji Varshitha
AI20BTECH11009

## Introduction

KNN algorithm, also known as Lazy learning is a non parametric model which is used to solve classification and regression problems.

## Algorithm

It follows the instance based learning as the learning part involves storing all the data points and performs the action at the time of classification or regression. Choosing the metric and value of K is very crucial in this model.

**Algorithm 1** KNN Algorithm

Calculate distance of new point from all the training data points.
Sort the distances in increasing order with corresponding train data and select first K data points.
**if** Classification **then**
    Class is determined by the majority class of the selected K data points.
**end if**
**if** Regression **then**
    The predicted continuous target value is given by mean or median value of the K data points.
**end if**

## How to determine K ?

Without loss of generality, K needs to be odd number to avoid ties in binary classification. For small values of K, the model has high variance and overfits the data. For example K = 1, the model is highly sensitive to outliers. Training error is low whereas test error is high.
For high values of K, model has less variance and increased bias.
We can determine optimal value of K by iterating through various values of K and choosing the one with minimum error rate.

## Distance metrics

Euclidean distance is the most common metric used. Besides, there are Manhattan, Minkowski and hamming distance metrics used when needed.