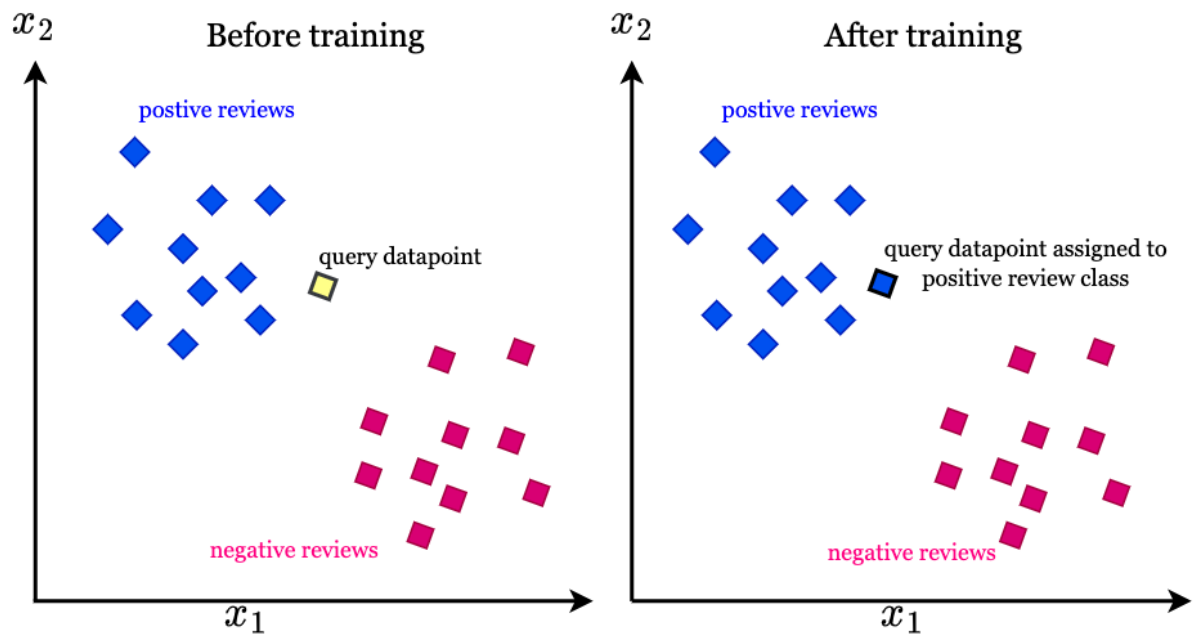




# OPDRACHT 5 – K NEAREST NEIGHBOURS



29 MEI 2024

Naam: Aymane Machrouki en Berke Ozmut  
Docent: Nick Goris

## Inhoudsopgave

Call tree.....	2
Totaal-uitleg.....	2
Resultaten .....	2
Antwoorden .....	3

## Call tree

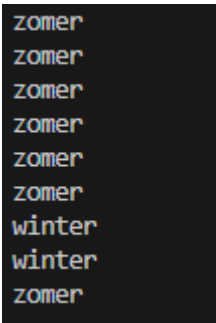
Main (Ik weet niet zeker of dit is wat er bedoelt word met call tree, maar ik dacht alle functies die je aanroept in de main.)

```
.....Normalize_datasets()
.....Create_labels_seasons()
.....Create_dataset_days
.....Best_k()
.....K_nearest_neighbours()
```

## Totaal-uitleg

Eerst worden alle benodigde datasets genormaliseerd met behulp van de functie **normalize\_datasets()**. Daarna worden alle benodigde labels aangemaakt met behulp van de functie **create\_labels\_seasons()**. Daarna wordt de dataset aangemaakt voor de dagen die wordt gebruikt voor het testen: **create\_dataset\_days()**. Nadat alle voorbereiding is voltooid wordt de functie **best\_k()** aangeroepen. Deze functie berekent de beste k in een bepaalde range en geeft de error percentage van de beste k en alle andere k's. Dit wordt gedaan door middel van de functie **k\_nearest\_neighbours()**. **K\_nearest\_neighbours()** telt de meest voorkomende neighbour. Dit wordt gedaan met behulp van de euclidean distance formule. De meest voorkomende neighbour (label) wordt vervolgens gebruikt als voorspelde label. Elke keer als er fout wordt voorspelt wordt dit, dus opgeteld en wordt hiervan een error percentage van weergegeven.

## Resultaten



```
zomer
zomer
zomer
zomer
zomer
zomer
winter
winter
zomer
```

## Antwoorden

What is the best value for k: `De 4e K heeft de laagste error percentage: 77.0%`

Also use your classifier to determine the season of the days:

```
zomer  
zomer  
zomer  
zomer  
zomer  
zomer  
winter  
winter  
zomer
```