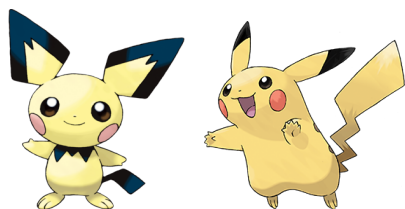


# Labb 1

**Syftet** med den här laborationen är att använda verktygen du lärt dig i Python för att implementera en förenklad maskininlärningsalgoritm.

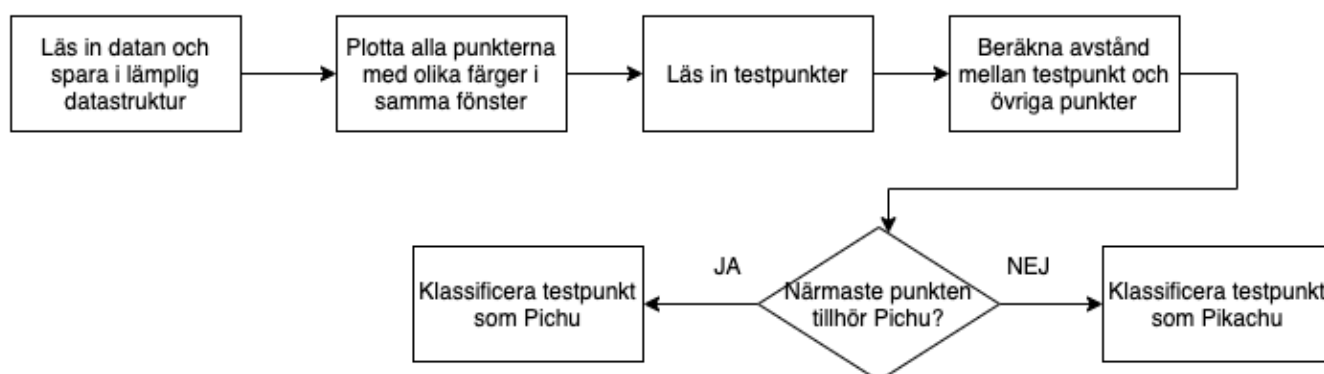


I den här laborationen finns (simulerad) data på Pichus och Pikachus längder och bredder. Du ska skapa en algoritm som baserat på den givna datan kunna avgöra om en ny data ska klassificeras som Pichu eller Pikachu.

**Notera:** du får inte använda databehandlingsbibliotek som pandas eller dylikt till denna uppgiften. Du får inte heller använda maskininlärningsbibliotek. Denna laboration syftar till att öva dig i Python, numpy och matplotlib.

## Grunduppgift

Följ detta flödesschema för att bygga den grundläggande algoritmen



Facit på givna testdatan:

```
Sample with (width, height): (25, 35) classified as Pikachu
Sample with (width, height): (15, 14) classified as Pichu
Sample with (width, height): (26, 30) classified as Pichu
Sample with (width, height): (22, 45) classified as Pikachu
```

## Uppgifter

Gör dessa uppgifter när du klassificerat testdatan korrekt.

1. Låt användaren mata in en testpunkt och avgör om den dess klass. Ta med felhanteringen som tar hand om negativa tal och icke-numeriska inputs. Se till att ha användarvänliga felmeddelanden.

2. Den approachen vi använt med närmaste punkten kan klassificera fel när punkterna för respektive klass går in i varandra. Nu ska du istället välja de fem närmaste punkterna till din testpunkt. Den klass testpunkten tillhör avgörs av majoritetsklassen av de närmaste punkterna.

### Bonusuppgifter (frivilliga)

3. Dela in ursprungsdatan slumpmässigt så att:
- 90 är träningsdata (45 Pikachu, 45 Pichu)
  - 10 är testdata (5 Pikachu, 5 Pichu)
4. Beräkna noggrannheten genom följande formel:

$$\text{accuracy} = \frac{\#TP + \#TN}{\text{Total}}$$

där

	Pikachu actual	Pichu actual
Pikachu predicted	TP	FP
Pichu predicted	FN	TN

Här har vi låtit Pikachu vara positiv och Pichu vara "icke-Pikachu" dvs negativ.

## Bedömning

Om du har fått någon kodsnuitt från någon annan eller hittat i någon sida är det **viktigt** att du källhänvisar. Detta gäller även om du fått från LLM som ChatGPT eller liknande. Skriv en kommentar bredvid koden som du har tagit.

### Godkänt

- löst grunduppgift, samt övriga uppgifter på korrekt sätt
- koden är kommenterad med relevanta kommentarer
- variabelnamnen är bra valda
- gjort flera relevanta git commits med beskrivande git-meddelanden

### Väl Godkänt

- koden är effektiv och enkel att följa
- koden är välstrukturerad med lämpliga funktioner och/eller OOP
- hög grad av DRY (don't repeat yourself) är uppnått
- kommentarerna är datavetenskapligt korrekta
- gjort samtliga uppgifter
- uppgifterna är lösta på högre nivå än för godkänt

