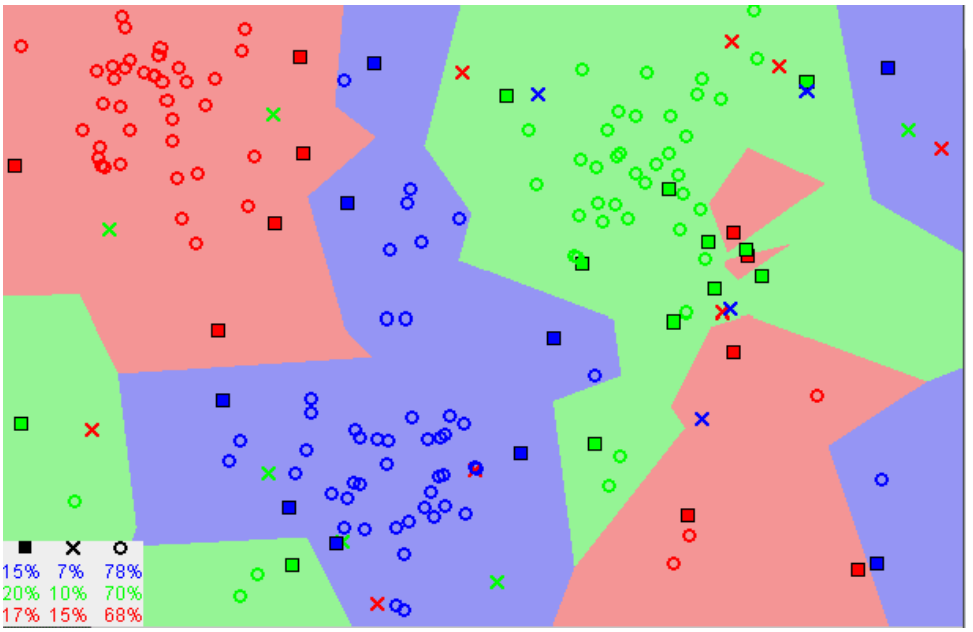
K-NEAREST NEIGHBOUR : WEEK 1

- The KNN algorithm assumes that similar things exist in close proximity. That is, Similar things are near to each other.



- KNN captures the idea of similarity of objects (distance, proximity, closeness) by calculating the distance between two points
- **Euclidean distance = √ a^2 + b^2**
- KNN works by finding the distances between a query and all the examples in the data, selecting the specified number of examples (K) closest to the query, then voting for the most frequent label (in the case of classification) or averages the labels (in the case of regression).

Why KNN might not work well?

- Run significantly slowly with large datasets

Applications?

- Movie Recommender system

PROJECT : MOVIE RECOMMENDER

IMPLEMENTATION DETAILS

- Database : Movie Lens Dataset from the University of Michigan
- PREPROCESSING
  - Converting all the movies not given ratings by a particular user to zero value ratings
  - Removing all users who have not rated up to 50 movies as they could not be legitimate
  - Removing all movies that have not received up to 10 ratings
  - Removing sparse data : Majority of the data is not rated

- MODEL :
  - KNN model with the following parameters
    - Number of neighbours : 20
    - algorithm : brute force (Calculate the euclidean distance from the query point to all the points in the dataset. Then we take the class with the majority points

- TOOLS / PACKAGES USED
  - Pandas (Loading data, preprocessing)
  - Matplotlib (plots)
  - Pickle (Saving model)
  - Streamlit
  - Jupyter Notebook
  - Vscode
  - Anaconda

- WORKING WITH PICKLE
  - Pickle is a python package that allows saving machine learning models

- STREAMLIT
  - import streamlit as st

- streamlit run app app.py
- Streamlit is a cloud platform that allows users to build and deploy machine learning apps
- I used streamlit to build an app for the movie recommender system
  - HOW THE APP WORKS:
    - Enter the name of a movie
    - Select the number of movie recommendations you would like to see using a slider
    - Get your recommendations

NEXT UP :

- Deploy the movie recommender system
- Add more data to make it robust
- Configure the Streamlit interface