# Question 3 Vector space Model and feature representation

Experiment with different representation techniques. Document your findings and make conclusions. Show how choosing n-gram features can influence your results

1. Representation techniques used
2. Findings (results)
3. Using n-grams (1, 2, 3, 4)
4. How n-grams influence results

For this segment of the course work we aimed to experiment with **count-based** representations. Word context or semantics were not considered. For the count-based representations, we utilized **term frequency** and **inverse-document term frequency** to evaluate the relative importance of each word or feature.

We represented each review as a matrix of **token counts** through the process called CountVectorization.

<https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html>

based on the **frequency** of a given term we represented