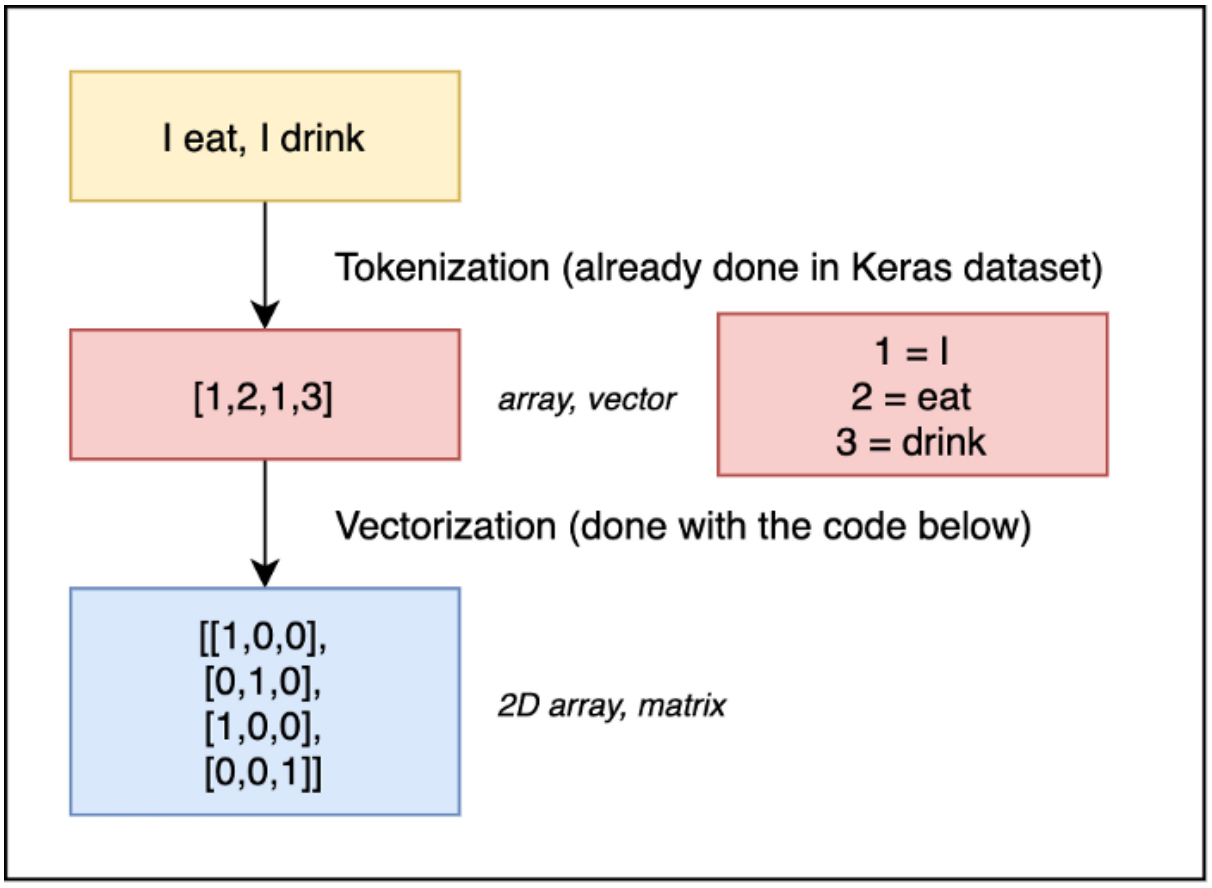
BINARY CLASSIFICATION USING MACHINE LEARNING / DEEP LEARNING

GOAL: Train a model that can predict the review by analysing the sentiment in the comment text.

DATASET: IMDB dataset of highly polar movie reviews in the form of text comments on different movies and a positive or negative rating. Comprises 25,000 movie reviews from IMDB, labelled by sentiment.

DATA IMPORT: Loaded directly from Keras (process_dataset.py)

DATA VECTORIZATION AND ONE-HOT ENCODING:



RESULTS:

Algorithm	Time (s)	Train Score	Test Score	Overfitting
Naive Bayes	1.14504	0.86884	0.83936	0.02948
Logistic Regression	136.041	0.98984	0.862	0.12784
K-Nearest Neighbor	0.156363	0.7854	0.62468	0.16072
Support Vector Machine	0.90942	0.86272	0.8514	0.01132
Decision Tree	80.1544	1	0.70808	0.29192
Random Forest	4.56928	0.8326	0.80008	0.03252
Voting Classifier	49.8875	0.91396	0.86408	0.04988

REFERENCE:

https://medium.com/thinkport/top-10-binary-classification-algorithms-a-beginners-guide-feeacbd7a3e2