Putting it all together

Over all, you start with a given text, you perform preprocessing, then you do feature extraction to convert text into numerical representation as follows:

Diagram, schematic

Description automatically generated

Your *X* becomes of dimension (m,3)(*m*,3) as follows.

Diagram

Description automatically generated

When implementing it with code, it becomes as follows:

Graphical user interface, text, application

Description automatically generated

You can see in the last step you are storing the extracted features as rows in your *X* matrix and you have *m* of these examples.