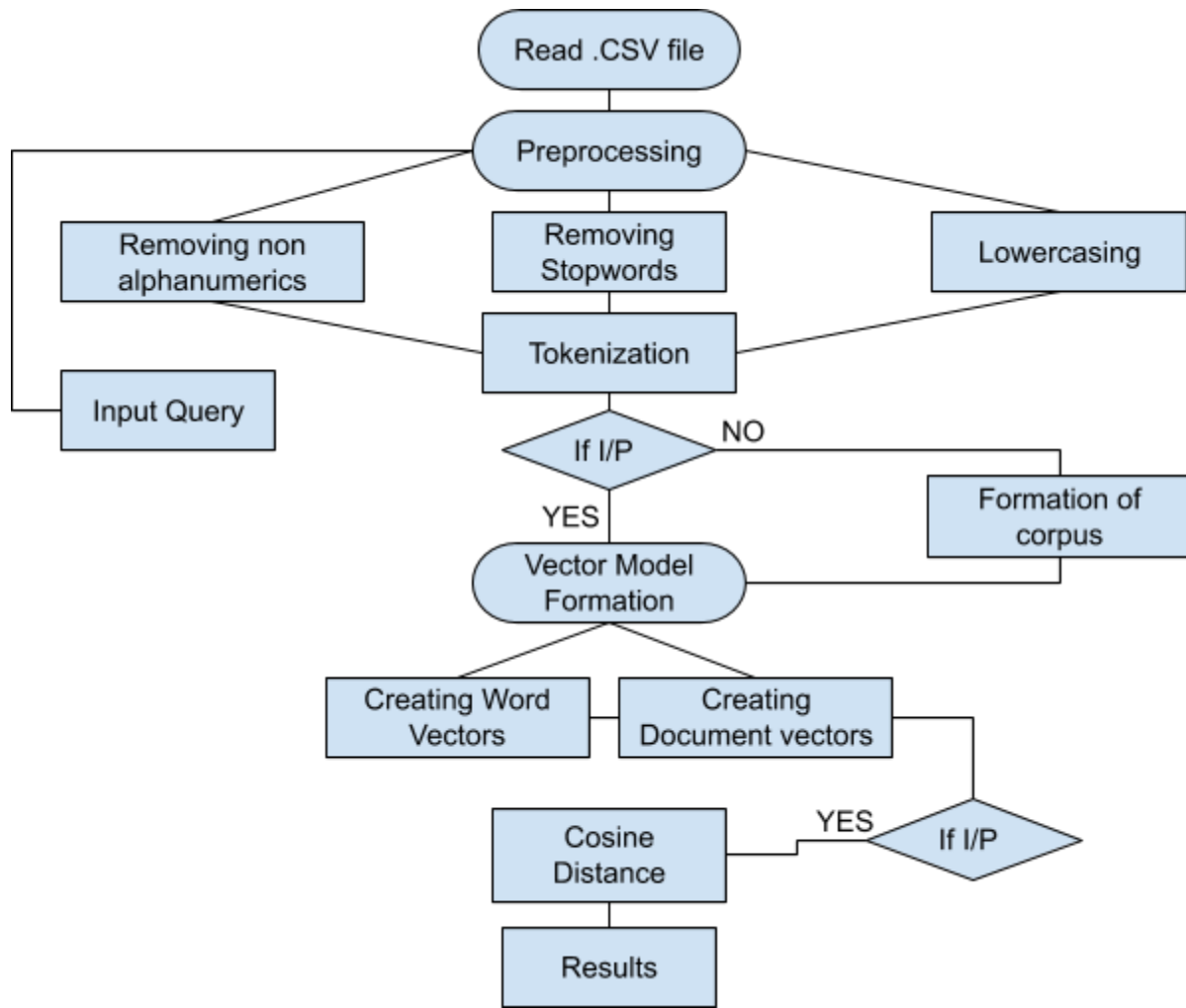


Design Documentation



Flowchart

The data structures Pandas DataFrame has been used to load the CSV file for performing further preprocessing. The list data structure has been used to store the tokenized words of each document and further each list formed is again stored as a list of lists.

Runtime for creating model - Approx 3 minutes

Runtime for returning query results - Approx 20 seconds