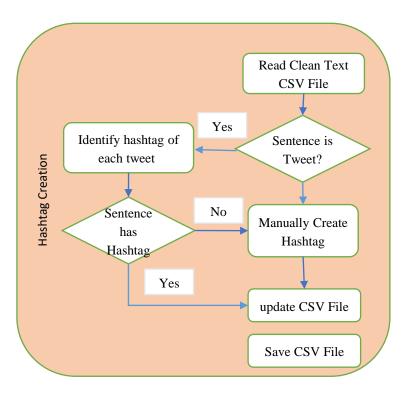


Share CSV file with 2+ Hindi
Language expert

Expert Labelling for each
Sentecne

Final Labelling based on
Voting



Remove extra space, RT, http / www links, "@" account text information, line break

Remove over expression (extra emoticons),

Remove greeting related msg like #सुप्रभात, #राम राम जी, जय श्रीराम etc.

Remove national or religious flag information like ININ

Remove non Hindi sentences

Remove sentences which are less than 4 words.

Remove duplicate sentences

Replace Hindi | (Hindi sentence marker with ".")

Clean Text