

Challenges :

- 1) Handling the punctuation symbols correctly and proper tokenisation for handling the excessive punctuation like !!!!!
- 2) Handling the ellipses at the end of the tweet.
- 3) Handling the different types of emoticons.
- 4) Handling the consecutive mentions like @abcd@xyz
- 5) Handling the consecutive hash-tags like #abcd#xyz
- 6) Differentiating between @ in e-mail and @ in mentions
- 7) fine differences like @abcd_ : is different from @abcd_
- 8) Handling url like tokens ex: https... where the last 3 dots are displaying as unicode and are wrong being classified as urls.