

Data preparation:

Data frame:

Body	Subject	word-Freq Body (1)	word-Freq Sub (2)	Date	Hour- (3)	minute

→ feature to be added

→ count of top 5 words in Phish body (4) ✓

→ count of top 5 words in Legit body (5) ✓

→ Day of the week (6) ✓

→ Count of capital words in Sub (7) ✓

→ Count Count Occurent of punctio ✓

→ word freq without Stop words ✓

→ Count vect results in higher dimension

→ reduce the feature from Count vec by taking only top 1000 words.

