

{ Overview: }

Input: Daily Drilling Report      Output: Segment name (20 segments)

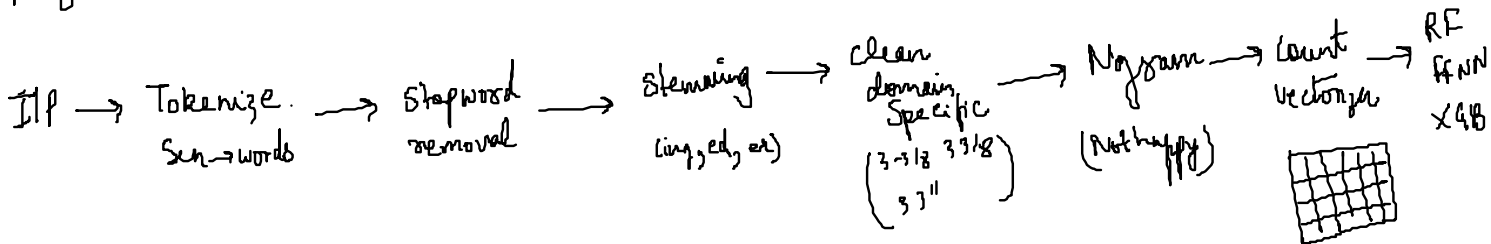
Example:

	Description	Activity
6943	circulated for survey	CIRC
7840	rig down casing running equipment, clean and c...	CASING/TUBING
7290	con't pooh 6 1/8" bha f/112m to 31m. l/d 8 3 1...	TRIPS
16655	cont. sweep hole till shaker clean.	CIRC
8410	recorded gyro from 20m to 3600m. pooh and laid...	OTHER - TOTCO/MWD/GYRO SURVEY
...	...	...
12621	washed up and reamed down last stand drilled. ...	CIRC
14742	connected test pump to kill line. p/t'd annula...	BOP AND WELLHEAD
20397	benon flush surface lines with wsw. pressure t...	CEMENTING
17311	drill 17-1/2"hole from 10m to 35m, rop: 25/6 =...	DRILL
814	m/up circulating swage/surface lines. circulat...	STUCK PIPE

Challenges:

1. Same matter is written in different styles by different people. ( tool got stuck, stuck tool, DLL MSFL Stuck)
2. Same word is written in different way by different people (POOH Pull out of Hole)
3. Large % of matter is not related to the segment. ( stop words, news, safety information)
4. Mix of numbers, abbreviations, words, joined words ( RIH, 9-3/8, drillpipe, drill pipe, pipe)
5. Unbalanced data

Algorithm.



Not applied: Pos Tag, Lemmatize, TF-IDF  
: Deep learning methods.

Results: Accuracy F2: 76%.