NLP Interestship

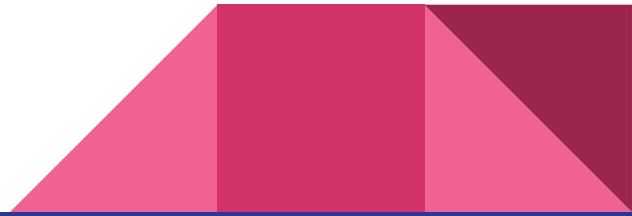
Module 4 : Exploratory Data Analysis (EDA), Logistic Regression, Naive Bayes, Confusion Matrix, TF-IDF, etc.

EDA

To be done by students as part of TASK-1.

Hints provided during NLP Session:

1. Find Question length,
2. Mean word length,
3. Word cloud, etc.



Logistic Regression

https://youtu.be/f6PNgEbopfQ

Naive Bayes

https://youtu.be/Pv75s9I8-mA

Performance Metrics

https://youtu.be/Fb6dXbDjCA8

TF-IDF

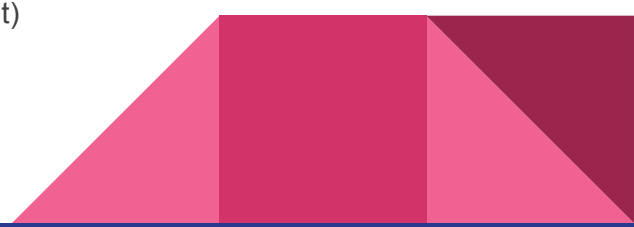
Term Frequency also known as TF measures the number of times a term (word) occurs in a document.

Normalized term frequency = no. of times a word occurs / total number_of_terms in the that doc.

The terms that are occurring more frequently doesn't always mean they are more important. Sometimes more frequently occurring terms are "the", "a","an", so we need some mechanism to weigh down the effect of too frequently occurring terms.

Inverse document frequency

IDF(word) = 1 + $\log_e$(Total Number Of Documents / Number Of Documents with that word in it)



THANK YOU