This pdf gives a brief idea (method) on which I have worked on the **2nd part**…  
Im explaining the code which i have written in trial.ipynb

**1**.First, I have analyzed the data using pandas and matplotlib by drawing some histograms on reviews.

**2**.Then I have started the main part which is preprocessing the text which is required for converting the text into vectors…

**3.PREPROCESSING:**

* Text is converted into lower case.
* Extra spaces and unnecessary new lines are removed.
* Punctuations are removed.(ALL THESE ARE DONE USING “re” library).
* Then the sentences are tokenized and the **stopwords** are removed.

**4.VECTORIZING:**

* Using TF-IDF Vectorizer we have converted the words into vectors and consider only the top 5000 terms ordered by term frequency across the corpus.

**5**. I have considered the columns “helpful\_votes” and encoded part of “text” for model training .

**6.**Logistic Regression with max\_iter=1000 and test-train split method is used for model training.