***The procedure was divided into 5 major steps:***

1. ***TEXT PREPROCESSING:***
   1. *We dropped the columns where the title was empty.*
   2. *We dropped the columns where description and bullet points both were empty.*
   3. *Removed stop words, unknown characters, punctuations etc*
   4. *Merged 3 columns(Bullet points,Description and Title) into a single column containing a single string.*
2. ***OUTLIER DETECTION:***
   1. *We checked the data frame to get a description of the numerical data and found that 99th percentile had value in 10,000 while 100th percentile had value up to the power of 10^9.*
   2. *We found that maximum product length was an outlier so we set a threshold to 10^4 and removed the rest of the values.*
   3. *For lower outliers the minimum value was 1th =and value near the 5th percentile was near 100.*
   4. *Thus we removed the outliers based on quartiles*
3. ***GENERATING WORD EMBEDDINGS***
   1. *We used Word2vec to generate word embeddings from the combined string which includes pre-processed text from TITLE , DESCRIPTION , BULLET\_POINTS.*
4. ***COMBINING FEATURES***
   1. *We combined the 2 features(Numerical = Length and Text = Word embeddings generated by word2vec )*
5. ***TESTING AND TRAINING***
   1. *Finally we trained the data on 2 features and generated results.*
   2. *Later we tested our model on the trained model.*