**SENTENCES**

**Converting Text into Sentences**

**To-Text Conversion (If PDF, Extract the Text Data) and Creation of CSV Data from the Above.**

**COLLECTION OF Data MANUSCRIPTS and Books.**

**(Formats – Docx,Pdf)**

**PREROCESSING**

1. Removing Stop words
2. Stemming
3. Lemmatizing
4. Part of Speech Tagging
5. Chunking
6. Applying N grams

**VECTORIZATION**

**Word2vec Embedding**

**Normalization**

**Dimensionality Reduction(LSA,PCA,SVD)**

**OUTPUT (Base-Summary)**

**1.) TRANSFER LEARNING MODEL (Using Bert , Roberta , XLnet )**

**2.)Base implementation of Transformers from scratch using available data.**

**HYPER PARAMETER FINE TUNING**

**SUMMARY (Best Output)**