Complete road map to prepare NLP:

NLP: Natural Language Processing,

Mostly dealing with the text data to analyze and work upon it as required.

Eg: Improve user experience or engagement with particular application via comments or feedback form the users.

Libraries to use:

ML:

SpaCy

Natural Language Processing with NLTK

DL:

PyTorch

Keras

TensorFlow

Transformers:

BERT,

(HuggingFace Library) (eg: sentence classification, spamham) -> use cases

Library: skatrain

Road Map:

Text Preprocessing Level -1

Tokenization

Lemmatization, stemming, POS

StopWords

Text Preprocessing level-2 (text to vectors)

Bag of Words

TFIDF

Unigrams,

Bigrams, n-grams

One hot encoding

Text Preprocessing – (Advance ways to text to vectors)

Gensim

Word2vec

AvgWord2vec

Solve ML use cases:

Eg: sentiment classifier, spam ham classifier, doc clasifier

(naïve bayes, multi nominal naïve bayes of multi class)

Get the understanding of Artificial Neural Network:

ANN works, loss function, optimizers, gradient boosting, etc

Understanding RNN, LSTM,GRU:  
 Sequence of data.

Eg: timeseries data, language translation, chatbots, q&A session,

Text Preprocessing level 3: (to overcome RNN limitaions)

Word embeddings (convert each individual into independent vector in a dictionary)

Word2vec

Bidirectional LSTM RNN, Encoders, and Decoders (sequence to sequence), Attention Models:

Next level.

Eg: in Lstm, if I have long sentence, the output of one neuron is not that promising, so to get the more information of the previous word of that sequence and it has memory to store that for furether uses

2 lstms with forward and backward approach

Next : Encoders and decoders’

Transformers:

BERT ( variation of Transformer) \_. Convert sentence in to a vector.

Bottom Top Approach Of Learning

1.Text Preprocessing Level 1- Tokenization,Lemmatization,StopWords,POS

2.Text Preprocessing Level 2- Bag Of Words, TFIDF, Unigrams,Bigrams,n-grams

3.Text Preprocessing- Gensim,Word2vec,AvgWord2vec

4.Solve Machine Learning Usecases

5.Get the Understanding Of Artificial Neural Network

6.Understanding Recurrent Neural Networks, LSTM,GRU

7.Text Preprocessing Level 3- Word Embeddings, Word2vec

8.Bidirectional LSTM RNN, Encoders And Decoders, Attention Models

9.Transformers

10.BERT