

Natural Language Processing [NLP]

application

- ① spam classifier → importance not "
- ② Restaurant Review → Food is good/bad
- ③ Alexa, Google assistant →
- ④ chatbot/ Q&A
- ⑤ Sentiment Analysis → +ve/-ve

NLP → Roadmap

* Text Preprocessing → Tokenization, stop words, stemming

② Bag of words, TFIDF, word2vec

* ANN, RNN, LSTM, GRU

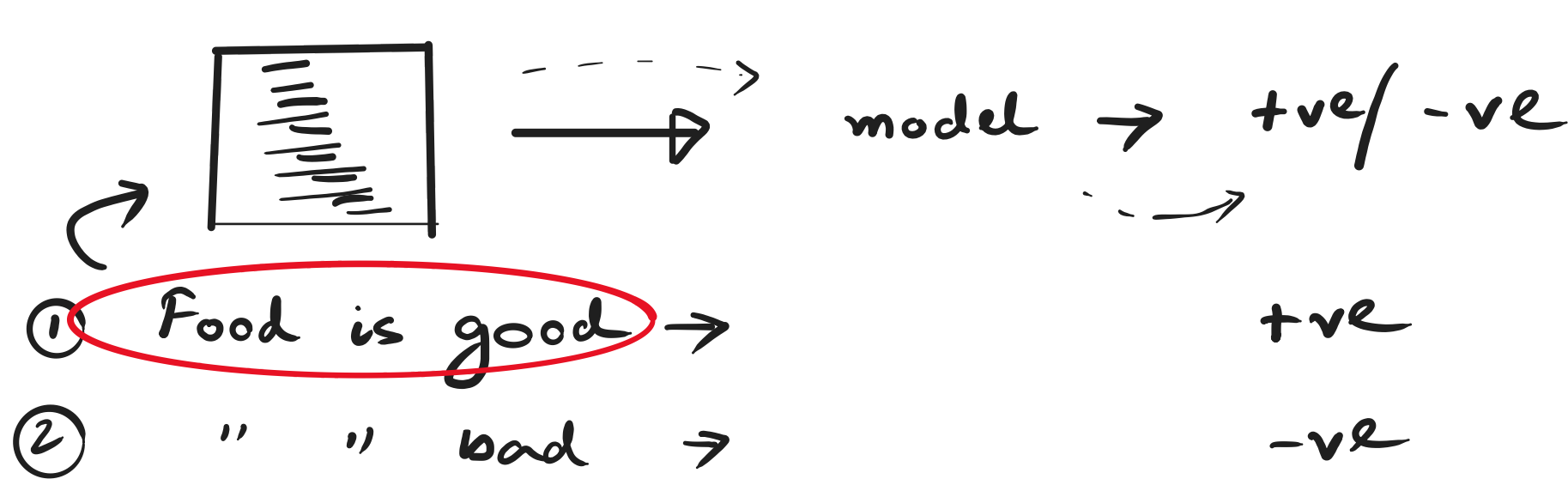
* seq2seq, attention, Encoders, decoders

* Transformers (GenAI)

* BERT, GPT

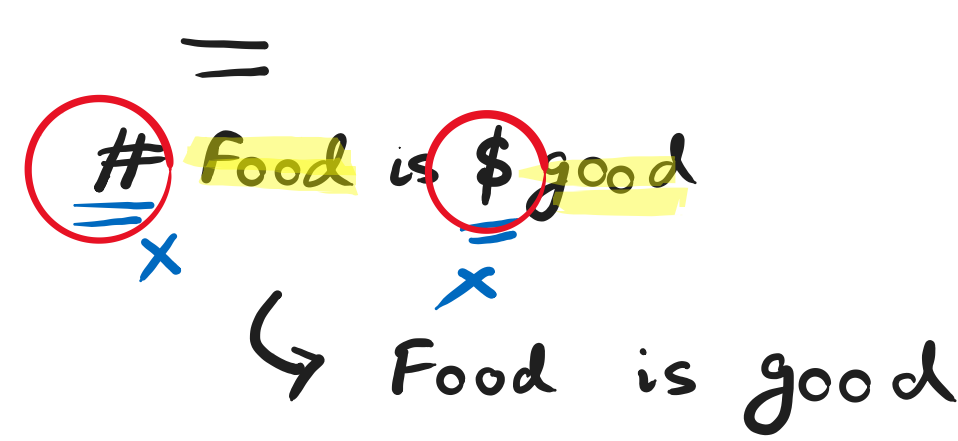
Text processing

words, sentences, paragraph, Document



----- → ----- (relevant/imp info)

Processing → ① Data cleaning ≡ ?, ..., #



② lowering (Lower case)

Food is not upto the mark but the food can be made good
 upper case lower case

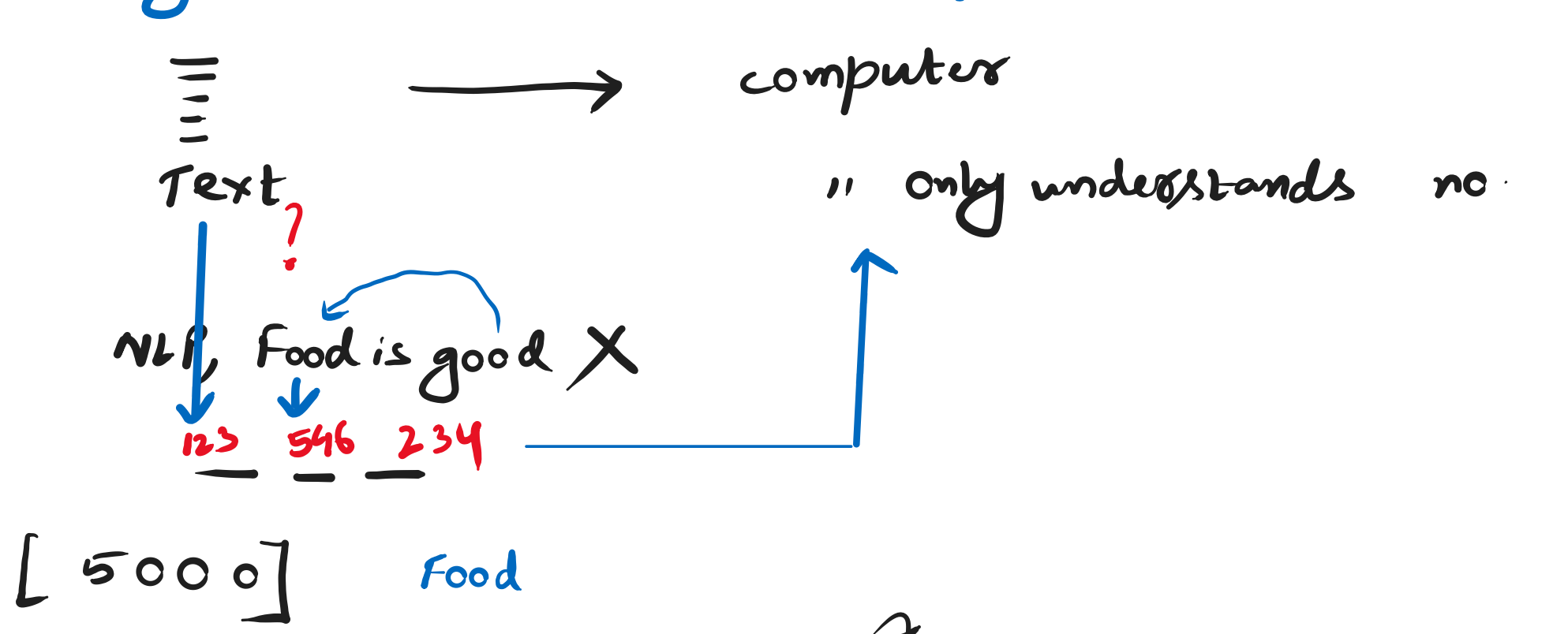
Tokenization → splitting sentences/paragraph into smaller units

eg Food isn't good → Food is not good
 Food is good → Food is good

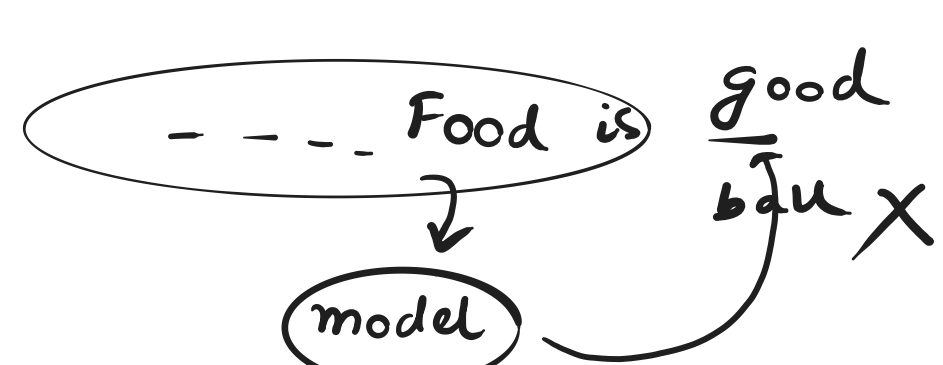
we are learning NLP. NLP has wide range of app. NLP is widely used now

we are

why Tokenization is needed?



-----? autocompletion
 next word prediction



Stopwords

I am a boy and I love to eat chicken → boy love eat chicken

stemming → Process of converting inflected word to stem/base word

history → histori
 historical → histori
 loving → love
 love → love
 final → fina
 finalization → fina
 finally → fina

Lemmatization → converts the words to base word with some meaning to it

eg history → history
 historical → history
 final → final
 finalization → final
 finally → final

Stemming

Lemmatization

- ① convert to base word
- ② Faster process
 historical → histori

eg Sentiment analysis

eg chatbot

