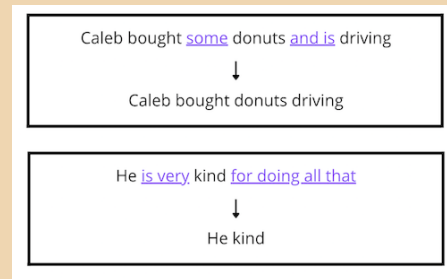ACM AI PRESENTS

RNNs in NLP

& Twitter Sentiment Analysis
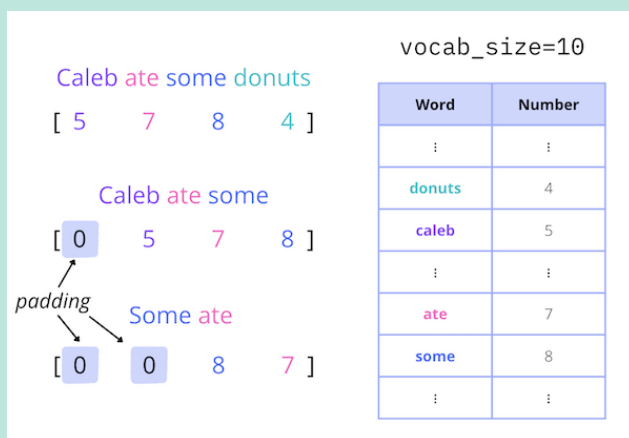
PREPROCESSING

Convert the data into a format acceptable for training the model
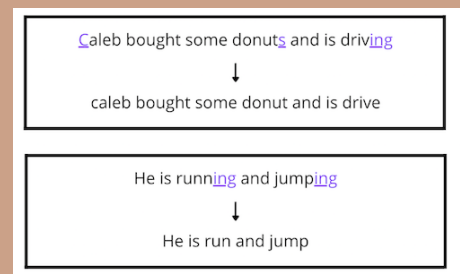
Caleb bought some donuts and is driving
↓
Caleb bought donuts driving

He is very kind for doing all that
↓
He kind

STOP WORDS

Words commonly used by humans that do not give much useful information

Caleb ate some donuts
[5   7   8   4]

Caleb ate some
[0   5   7   8]

padding

Some ate
[0   0   8   7]

vocab_size=10

Word	Number
⋮	⋮
donuts	4
caleb	5
⋮	⋮
ate	7
some	8
⋮	⋮

ONE-HOT ENCODING

Represents categorical variables as binary vectors

Caleb bought some donuts and is driving
↓
caleb bought some donut and is drive

He is running and jumping
↓
He is run and jump

STEMMING

Reducing a word to its stem by removing its affixes, helpful in preprocessing

Simple Recurrent Neural Net

i_t h_t O_t

i_1 h_1 O_1
i_2 h_2 O_2
i_n h_n O_n
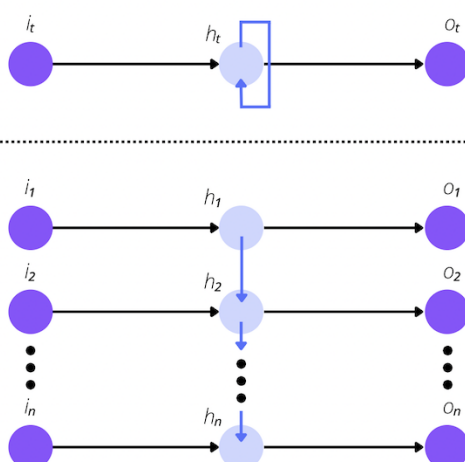
RECURRENT NEURAL NETWORKS

Process sequence input by iterating through the elements. RNNs pass the outputs from one timestep to their input on the next timestep