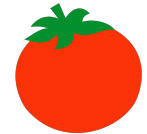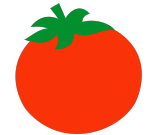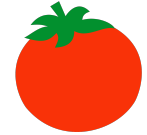# Activity 9:
Building Text Features

ML for Health, Week 9

ANSWER KEY

Dune (2021) Reviews

- A spicy journey to an amazing future

- Unlike anything we've seen in science fiction before

- A smart science fiction film

- [We've] seen too much of this sort of thing before

- An effort to create a grand science fiction epic has yielded something cold

Part 1: Define the vocabulary
(feel free to discard common words like 'a' and 'to')

- A spicy journey to an amazing future

- Unlike anything we've seen in science fiction before

- A smart science fiction film

- [We've] seen too much of this sort of thing before

- An effort to create a grand science fiction epic has yielded something cold

spicy
journey
amazing
future
unlike
anything
seen
science
fiction
before
smart
film
much
sort
thing
effort
create
grand
epic
yielded
something
cold

Part 2: Use your vocabulary to create feature vectors for <u>two</u> of the reviews

- A spicy journey to an amazing future

- Unlike anything we've seen in science fiction before

- A smart science fiction film

- [We've] seen too much of this sort of thing before

- An effort to create a grand science fiction epic has yielded something cold

spicy	journey	amazing	future	unlike	anything	seen	science	fiction	before	smart	film	much	sort	thing	effort	create	grand	epic	yielded	something	cold
1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1

Part 3: Calculate tf-idf features for <u>two</u> words from different reviews

- A **spicy** journey to an amazing future

- Unlike anything we've seen in science fiction before

- A smart **science** fiction film

- [We've] seen too much of this sort of thing before

- An effort to create a grand science fiction epic has yielded something cold

SPICY:

tf = 1/7
df = 1/5
tf-idf = 5/7

SCIENCE:

tf = 1/5
df = 3/5
tf-idf = 1/3