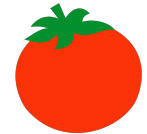
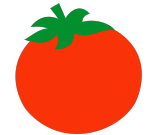
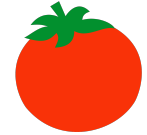


# Activity 9: Building Text Features

ML for Health, Week 9

# 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')

write  
vocab  
words  
here

- 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 2: Use your vocabulary to create feature vectors for two 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

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

write  
vocab  
words  
here

## Part 3: Calculate tf-idf features for two 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 =

df =

tf-idf =

SCIENCE:

tf =

df =

tf-idf =