Please find the link to he IMDB reviews dataset [here](http://ai.stanford.edu/~amaas/data/sentiment/)

You will find here 50,000 movie reviews which are classified as positive of negative.

<http://projector.tensorflow.org/>

Please find [here](https://github.com/tensorflow/datasets/tree/master/docs/catalog) the datasets GitHub url.

For more information, please checkout the [TensorFlow datasets documentation](https://www.tensorflow.org/datasets/catalog/overview).

The URL shown in the video is outdated. Please use the updated URL shown below:

<https://github.com/tensorflow/datasets/tree/master/docs/catalog>

The URL shown in the video is outdated. Please use the updated URL shown below:

<https://github.com/tensorflow/datasets/blob/master/docs/catalog/imdb_reviews.md>

**This week you looked at taking your tokenized words and using Embeddings to establish meaning from them in a mathematical way. Words were mapped to vectors in higher dimensional space, and the semantics of the words then learned when those words were labelled with similar meaning. So, for example, when looking at movie reviews, those movies with positive sentiment had the dimensionality of their words ending up ‘pointing’ a particular way, and those with negative sentiment pointing in a different direction. From this, the words in future sentences could have their ‘direction’ established, and from this the sentiment inferred. You then looked at sub word tokenization, and saw that not only do the meanings of the words matter, but also the sequence in which they are found.**

**You’ll look into that next week.**