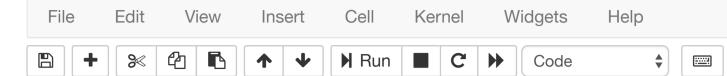
Not Trusted

Python 3 O



## Text classification with movie reviews

This notebook classifies movie reviews as *positive* or *negative* using the text of the review. This is an example of *binary*—or two-class—classification, an important and widely applicable kind of machine learning problem.

We'll use the <u>IMDB dataset</u> that contains the text of 50,000 movie reviews from the <u>Internet Movie Database</u>. These are split into 25,000 reviews for training and 25,000 reviews for testing. The training and testing sets are *balanced*, meaning they contain an equal number of positive and negative reviews.

This notebook uses <u>tf.keras</u>, a high-level API to build and train models in TensorFlow. For a more advanced text classification tutorial using <u>tf.keras</u>, see the MLCC Text Classification Guide.

```
In [5]: import tensorflow as tf
    from tensorflow import keras
    from google.colab import files

import numpy as np

print(tf.__version__)

1.12.0-rc1
```

## **Download the IMDB dataset**

The IMDB dataset comes packaged with TensorFlow. It has already been preprocessed such that the reviews (sequences of words) have been converted to sequences of integers, where each integer represents a specific word in a dictionary.

The following code downloads the IMDB dataset to your machine (or uses a cached copy if you've already downloaded it):