Writing TFRecords

Write serialized Example objects into TFRecords files.

Chapter Goals:

- Write the training and evaluation set data into TFRecords files

A. Writing Example data

Now that we've completed the function to convert each DataFrame row into an Example object, we can create the efficient input pipeline storage for both the training and evaluation sets. The data storage will be in the form of TFRecords files, which hold serialized Example objects.

The `write_tfrecords` function (shown below) writes the data from a given DataFrame into a TFRecords file. It uses the `create_example` function from the previous chapter to convert each row of the dataset into an Example object. Each Example object is then serialized and written into the TFRecords file.

```python
import tensorflow as tf

# Write serialized Example objects to a TFRecords file
def write_tfrecords(dataset, has_labels, tfrecords_file):
    writer = tf.python_io.TFRecordWriter(tfrecords_file)
    for i in range(len(dataset)):
        example = create_example(dataset.iloc[i], has_labels)
        writer.write(example.SerializeToString())
    writer.close()
```

We can use the above function to write the training set's serialized Example data into a TFRecords file called *train.tfrecords* and the evaluation set's serialized Example data into a TFRecords file called *eval.tfrecords*. These files will then be used in the input pipeline for the machine learning model.

```python
# train_set is the training DataFrame
write_tfrecords(train_set, 'train.tfrecords')

# eval_set is the evaluation DataFrame
write_tfrecords(eval_set, 'eval.tfrecords')
```