```
B211038- Srushti Gavale

IMDB dataset
```

In [1]:

```
import numpy as np
from tensorflow.keras.datasets import imdb
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout, Embedding, Flatten
from tensorflow.keras.preprocessing.sequence import pad_sequences

# Load the IMDB dataset
(x_train, y_train), (x_test, y_test) = imdb.load_data(num_words=10000)
```

```
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/imdb.npz (http
s://storage.googleapis.com/tensorflow/tf-keras-datasets/imdb.npz)
17464789/17464789 [==============================] - 0s 0us/step
```

In [2]:

```
max_len = 500

# Pad and truncate the sequences
x_train = pad_sequences(x_train, maxlen=max_len)
x_test = pad_sequences(x_test, maxlen=max_len)
```

In [3]:

```
model = Sequential()
model.add(Embedding(10000, 32, input_length=max_len))
model.add(Flatten())
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(1, activation='sigmoid'))

model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
```

In [4]:

```
model.fit(x_train, y_train, validation_split=0.2, epochs=5, batch_size=128)
```

```
Epoch 1/5
157/157 [==============================] - 21s 124ms/step - loss: 0.5079 - accuracy: 0.7199 - val
_loss: 0.2970 - val_accuracy: 0.8758
Epoch 2/5
157/157 [==============================] - 17s 108ms/step - loss: 0.1730 - accuracy: 0.9354 - val
_loss: 0.3071 - val_accuracy: 0.8746
Epoch 3/5
157/157 [==============================] - 17s 111ms/step - loss: 0.0490 - accuracy: 0.9876 - val
_loss: 0.3857 - val_accuracy: 0.8654
Epoch 4/5
157/157 [==============================] - 13s 81ms/step - loss: 0.0111 - accuracy: 0.9990 - val_
loss: 0.4101 - val_accuracy: 0.8720
Epoch 5/5
157/157 [==============================] - 12s 78ms/step - loss: 0.0035 - accuracy: 0.9999 - val_
loss: 0.4421 - val_accuracy: 0.8744
```

Out[4]:

```
<keras.callbacks.History at 0x7f2c428b6100>
```

In [5]:

```
loss, accuracy = model.evaluate(x_test, y_test)
print(f'Test accuracy: {accuracy * 100:.2f}%')
```

```
782/782 [==============================] - 5s 6ms/step - loss: 0.4469 - accuracy: 0.8720
Test accuracy: 87.20%
```

In [6]:

```
def predict_review(review):
    # Convert the review to a sequence of word indices
    seq = imdb.get_word_index()
    words = review.split()
    seq = [seq[w] if w in seq else 0 for w in words]
    seq = pad_sequences([seq], maxlen=max_len)

    # Make the prediction
    pred = model.predict(seq)[0]

    # Return the prediction
    return 'positive' if pred >= 0.5 else 'negative'

review = "This movie was great! I loved the story and the acting was superb."
prediction = predict_review(review)
print(f'Review: {review}')
print(f'Prediction: {prediction}')
```

```
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/imdb_word_inde
x.json (https://storage.googleapis.com/tensorflow/tf-keras-datasets/imdb_word_index.json)
1641221/1641221 [==============================] - 0s 0us/step
1/1 [==============================] - 0s 163ms/step
Review: This movie was great! I loved the story and the acting was superb.
Prediction: positive
```

In [7]:

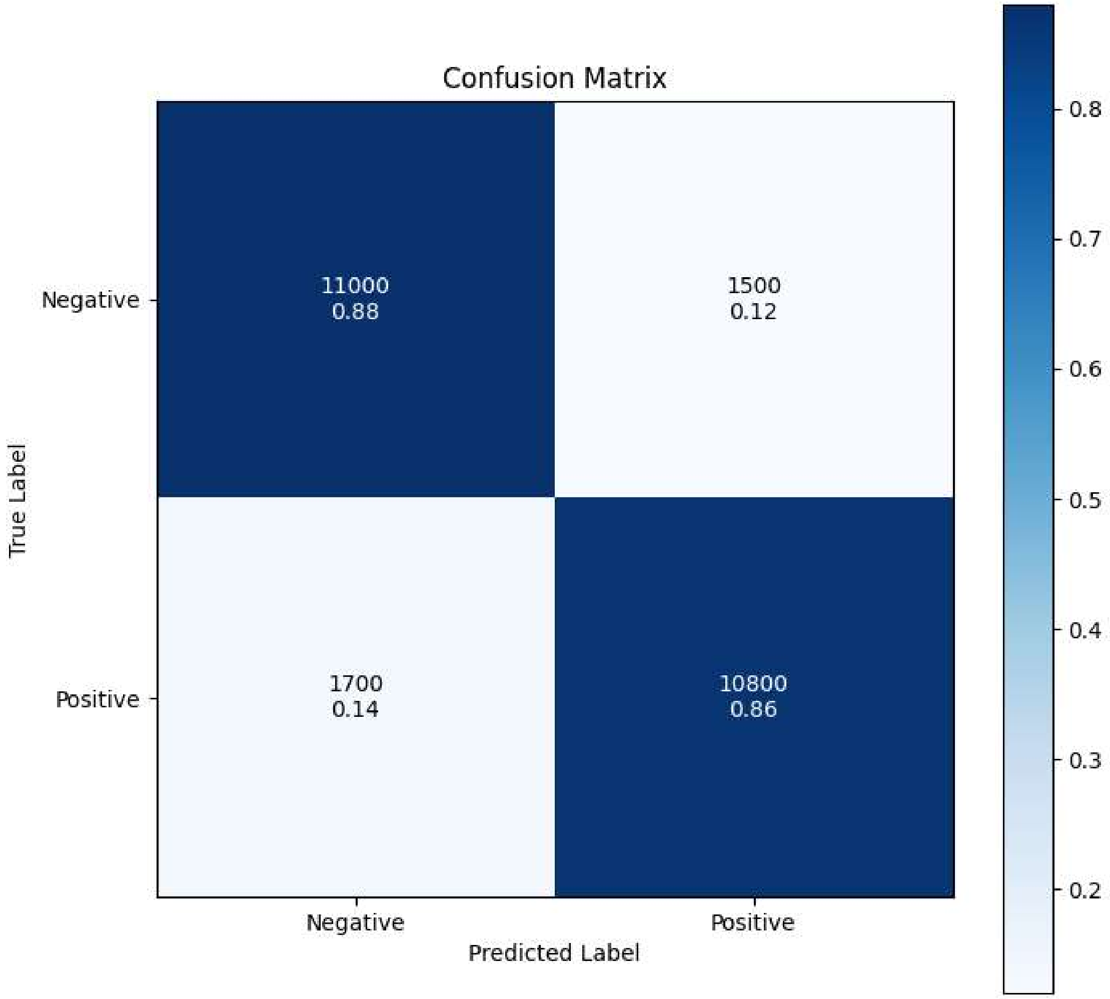
```
# Print model summary
model.summary()
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 500, 32)           320000

 flatten (Flatten)           (None, 16000)             0

 dense (Dense)               (None, 128)               2048128

 dropout (Dropout)           (None, 128)               0

 dense_1 (Dense)             (None, 1)                 129

=================================================================
Total params: 2,368,257
Trainable params: 2,368,257
Non-trainable params: 0
_________________________________________________________________
```

In [8]:

```
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt
import numpy as np

# Get predicted labels
y_pred = np.round(model.predict(x_test))

# Generate confusion matrix
cm = confusion_matrix(y_test, y_pred)

# Normalize confusion matrix
cm_norm = cm.astype('float') / cm.sum(axis=1)[:, np.newaxis]

# Set up plot
fig, ax = plt.subplots(figsize=(8, 8))

# Plot confusion matrix
im = ax.imshow(cm_norm, interpolation='nearest', cmap=plt.cm.Blues)
ax.figure.colorbar(im, ax=ax)

# Set labels
ax.set(xticks=np.arange(cm.shape[1]),
       yticks=np.arange(cm.shape[0]),
       xticklabels=['Negative', 'Positive'], yticklabels=['Negative', 'Positive'],
       title='Confusion Matrix',
       ylabel='True Label',
       xlabel='Predicted Label')

# Add labels to each cell
thresh = cm_norm.max() / 2.
for i in range(cm_norm.shape[0]):
    for j in range(cm_norm.shape[1]):
        ax.text(j, i, format(cm[i, j], 'd') + '\n' + format(cm_norm[i, j], '.2f'),
                ha="center", va="center",
                color="white" if cm_norm[i, j] > thresh else "black")

# Show plot
plt.show()
```

```
782/782 [==============================] - 6s 8ms/step
```



In [9]:

```
from sklearn.metrics import classification_report

print(classification_report(y_test, y_pred, target_names=['Negative', 'Positive']))
```

```
              precision    recall  f1-score   support

    Negative       0.87      0.88      0.87     12500
    Positive       0.88      0.86      0.87     12500

    accuracy                           0.87     25000
   macro avg       0.87      0.87      0.87     25000
weighted avg       0.87      0.87      0.87     25000
```

In [ ]: