

# Summer 2024: CS5720 NNDL - ICP-6

Lasya Vanga (700762893)

GitHub Link: <https://github.com/Lasya-vanga/NNDL-ICP6>

1. Save the model and use the saved model to predict on new text data (ex, “A lot of good things are happening. We are respected again throughout the world, and that's a great thing.@realDonaldTrump”)

Solution:

The image displays two screenshots of a Google Colab notebook titled "NNDL-ICP6-700762893.ipynb".

The first screenshot shows the initial setup code in a cell:

```
[3] !pip install tensorflow
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Embedding, LSTM, SpatialDropout1D
from matplotlib import pyplot
from sklearn.model_selection import train_test_split
from tensorflow.keras.utils import to_categorical # Updated import
import re

from sklearn.preprocessing import LabelEncoder

data = pd.read_csv('Sentiment.csv')
# Keeping only the necessary columns
data = data[['text', 'sentiment']]

data['text'] = data['text'].apply(lambda x: x.lower())
data['text'] = data['text'].apply(lambda x: re.sub('[^a-zA-z0-9\s]', '', x))

for idx, row in data.iterrows():
    row[0] = row[0].replace('\r', ' ')

max_features = 2000
tokenizer = Tokenizer(num_words=max_features, split=' ')
tokenizer.fit_on_texts(data['text'].values)
X = tokenizer.texts_to_sequences(data['text'].values)

X = pad_sequences(X)

embed_dim = 128
lstm_out = 196
```

The second screenshot shows the model definition and training code in another cell:

```
def createmodel():
    model = Sequential()
    model.add(Embedding(max_features, embed_dim, input_length = X.shape[1]))
    model.add(LSTM(lstm_out, dropout=0.2, recurrent_dropout=0.2))
    model.add(Dense(3, activation='softmax'))
    model.compile(loss = 'categorical_crossentropy', optimizer='adam', metrics = ['accuracy'])
    return model

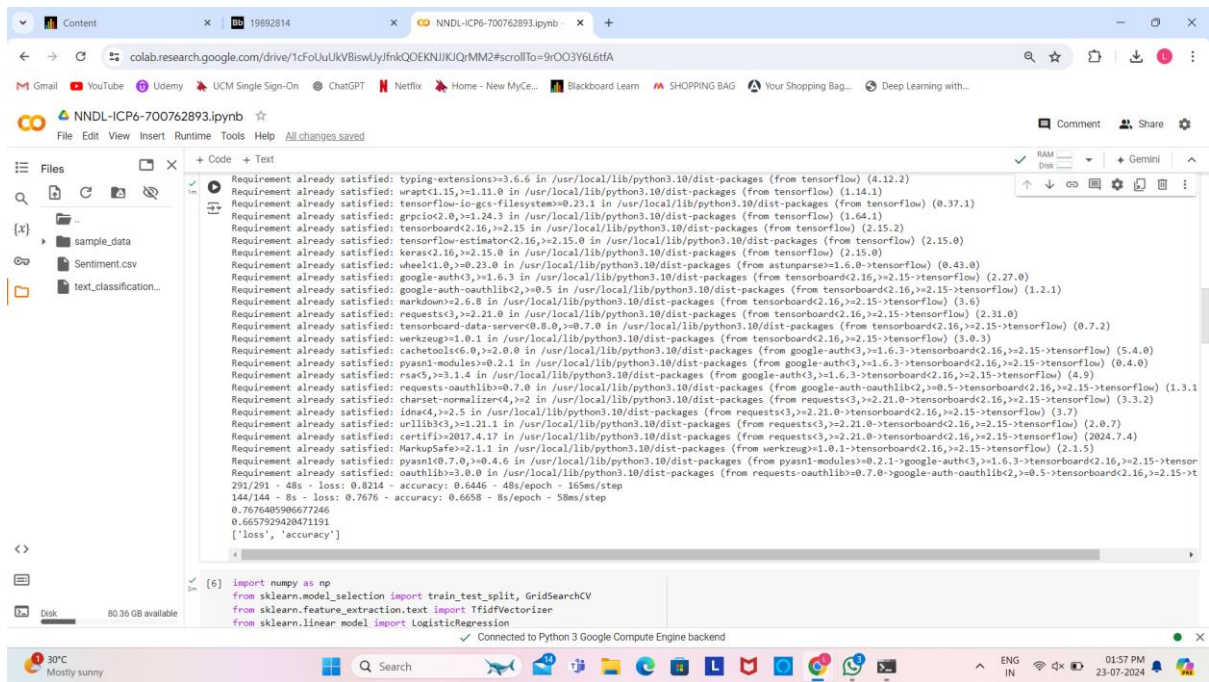
# print(model.summary())

labelencoder = LabelEncoder()
integer_encoded = labelencoder.fit_transform(data['sentiment'])
y = to_categorical(integer_encoded)
X_train, X_test, Y_train, Y_test = train_test_split(X, y, test_size = 0.33, random_state = 42)

batch_size = 32
model = createmodel()
model.fit(X_train, Y_train, epochs = 1, batch_size=batch_size, verbose = 2)
score, acc = model.evaluate(X_test, Y_test, verbose=2, batch_size=batch_size)
print(score)
print(acc)
print(model.metrics_names)
```

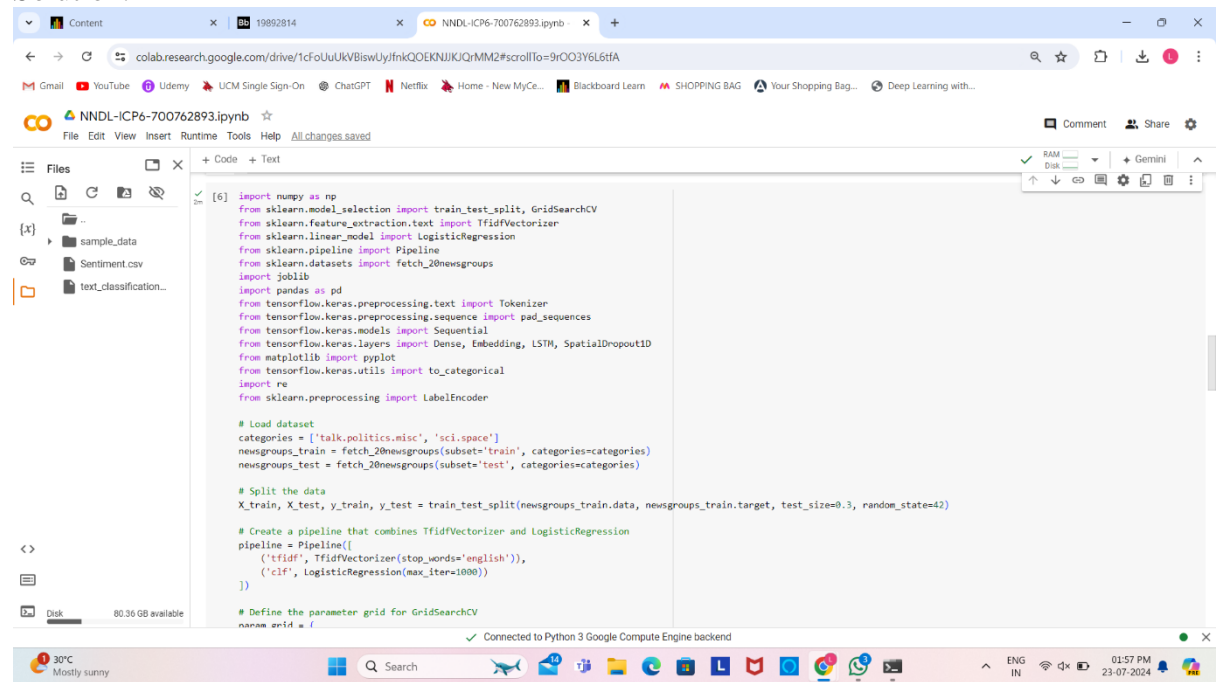
Below the code, the output of the model summary and evaluation is shown:

```
Requirement already satisfied: tensorflow in /usr/local/lib/python3.10/dist-packages (2.15.0)
Requirement already satisfied: absl-py>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.4.0)
Requirement already satisfied: astunparse>=1.6.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.6.3)
Requirement already satisfied: flatbuffers>=23.5.26 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (24.3.25)
Requirement already satisfied: gast>=0.5.0, <=0.5.1, <=0.5.2, >=0.2.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.6.0)
Requirement already satisfied: google-pasta>=0.1.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.2.0)
Requirement already satisfied: h5py>=2.9.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.9.0)
Requirement already satisfied: libclang>=13.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (18.1.1)
Requirement already satisfied: ml-dtypes==0.2.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.2.0)
Requirement already satisfied: numpy<2.0.0, >=1.23.5 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.25.2)
```



## 2. Apply GridSearchCV on the source code provided in the class

### Solution:



Content x 19892814 x NNDL-ICP6-700762893.ipynb x +

colab.research.google.com/drive/1cfoUuUkVBiswUyJfnkQOEKNJJKQrMM2#scrollTo=9rOO3Y6L6t6A

Gmail YouTube Udem UCM Single Sign-On ChatGPT Netflix Home - New MyCe... Blackboard Learn SHOPPING BAG Your Shopping Bag... Deep Learning with...

NNDL-ICP6-700762893.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

Files

sample\_data

Sentiment.csv

text\_classification...

```
[6] } 'cif_c': (0.1, 1, 10),

# Create a GridSearchCV object
grid_search = GridSearchCV(pipeline, param_grid, cv=5)

# Fit the GridSearchCV object to the training data
grid_search.fit(X_train, y_train)

# Print the best parameters found by GridSearchCV
print("Best parameters:", grid_search.best_params_)

# Get the best model from GridSearchCV
best_model = grid_search.best_estimator_

# Save the best model to a file
model_filename = 'text_classification_model.pkl'
joblib.dump(best_model, model_filename)

# Load the model from the file
loaded_model = joblib.load(model_filename)

# New text data for prediction
new_text = ["A lot of good things are happening. We are respected again throughout the world, and that's a great thing.@realDonaldTrump"]

# Make predictions
predictions = loaded_model.predict(new_text)
print(predictions)

!pip install tensorflow

data = pd.read_csv('Sentiment.csv')
# Keeping only the necessary columns
```

Connected to Python 3 Google Compute Engine backend

30°C Mostly sunny

Content x 19892814 x NNDL-ICP6-700762893.ipynb x +

colab.research.google.com/drive/1cfoUuUkVBiswUyJfnkQOEKNJJKQrMM2#scrollTo=9rOO3Y6L6t6A

Gmail YouTube Udem UCM Single Sign-On ChatGPT Netflix Home - New MyCe... Blackboard Learn SHOPPING BAG Your Shopping Bag... Deep Learning with...

NNDL-ICP6-700762893.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

Files

sample\_data

Sentiment.csv

text\_classification...

```
[6] for idx, row in data.iterrows():
    row[8] = row[8].replace('rt', ' ')

max_features = 2000
tokenizer = Tokenizer(num_words=max_features, split=' ')
tokenizer.fit_on_texts(data['text'].values)
X = tokenizer.texts_to_sequences(data['text'].values)
X = pad_sequences(X)

embed_dim = 128
lstm_out = 196
def create_model():
    model = Sequential()
    model.add(Embedding(max_features, embed_dim, input_length = X.shape[1]))
    model.add(LSTM(lstm_out, dropout=0.2, recurrent_dropout=0.2))
    model.add(Dense(3, activation='softmax'))
    model.compile(loss = 'categorical_crossentropy', optimizer='adam', metrics = ['accuracy'])
    return model

# print(model.summary())

labelencoder = LabelEncoder()
integer_encoded = labelencoder.fit_transform(data['sentiment'])
y = to_categorical(integer_encoded)
X_train, X_test, Y_train, Y_test = train_test_split(X, y, test_size = 0.33, random_state = 42)

batch_size = 32
model = create_model()
model.fit(X_train, Y_train, epochs = 1, batch_size=batch_size, verbose = 2)
score, acc = model.evaluate(X_test, Y_test, verbose=2, batch_size=batch_size)
print(score)
print(acc)
print(model.metrics_names)
```

Connected to Python 3 Google Compute Engine backend

30°C Mostly sunny

Content 19892814 NNDL-ICP6-700762893.ipynb

colab.research.google.com/drive/1cfoUuUKvBiswUyJfnkQOEKNJKQrMM2#scrollTo=9rOO3Y6L6tfa

Gmail YouTube Udem UCM Single Sign-On ChatGPT Netflix Home - New MyCe... Blackboard Learn SHOPPING BAG Your Shopping Bag... Deep Learning with...

NNDL-ICP6-700762893.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files + Code + Text

sample\_data Sentiment.csv text\_classification...

Best parameters: {'clf\_c': 10, 'tfidf\_max\_df': 0.5}

[6]

Requirement already satisfied: tensorflow in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.15.0)

Requirement already satisfied: absl-py>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.4.0)

Requirement already satisfied: astunparse>=1.6.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.6.3)

Requirement already satisfied: flatbuffers>=23.5.26 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (24.3.25)

Requirement already satisfied: gast!=0.5.0,!=0.5.1,!=0.5.2,>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.6.0)

Requirement already satisfied: google-pasta>=0.1.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.2.0)

Requirement already satisfied: h5py>=2.9.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.9.0)

Requirement already satisfied: libclang>=13.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (18.1.1)

Requirement already satisfied: ml-dtypes>=0.2.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.2.0)

Requirement already satisfied: numpy<2.0.0,>=1.23.5 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.25.2)

Requirement already satisfied: opt-einsum>=2.3.2 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.3.0)

Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from tensorflow) (24.1)

Requirement already satisfied: protobuf<4.21.0,!=4.21.1,!=4.21.2,!=4.21.3,!=4.21.4,!=4.21.5,<5.0.0dev,>=3.20.3 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.20)

Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from tensorflow) (67.7.2)

Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.16.0)

Requirement already satisfied: termcolor>=1.1.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.4.0)

Requirement already satisfied: typing-extensions>=3.6.6 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (4.12.2)

Requirement already satisfied: wrapt<1.15,>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.14.1)

Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.37.1)

Requirement already satisfied: grpcio<2.0,>=1.24.3 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.64.1)

Requirement already satisfied: tensorflow-estimator<2.16,>=2.15.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.15.2)

Requirement already satisfied: tensorflow-estimator<2.16,>=2.15.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.15.0)

Requirement already satisfied: keras<2.16,>=2.15.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.15.0)

Requirement already satisfied: wheel<1.0,>=0.23.0 in /usr/local/lib/python3.10/dist-packages (from astunparse>=1.6.0->tensorflow) (0.43.0)

Requirement already satisfied: google-auth<2.16,>=2.15 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.27.0)

Requirement already satisfied: google-auth-oauthlib<2,>=0.5 in /usr/local/lib/python3.10/dist-packages (from tensorflowboard<2.16,>=2.15->tensorflow) (1.2.1)

Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.10/dist-packages (from tensorflowboard<2.16,>=2.15->tensorflow) (3.6)

Requirement already satisfied: requests<3,>=2.21.0 in /usr/local/lib/python3.10/dist-packages (from tensorflowboard<2.16,>=2.15->tensorflow) (2.31.0)

Requirement already satisfied: tensorflow-data-server<0.8,>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from tensorflowboard<2.16,>=2.15->tensorflow) (0.7.2)

Requirement already satisfied: Werkzeug>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from tensorflowboard<2.16,>=2.15->tensorflow) (3.0.3)

Requirement already satisfied: cachetools<6.0,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorflowboard<2.16,>=2.15->tensorflow) (5.4.0)

Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorflowboard<2.16,>=2.15->tensorflow) (0.4.0)

Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorflowboard<2.16,>=2.15->tensorflow) (4.9)

Connected to Python 3 Google Compute Engine backend

30°C Mostly sunny

Content 19892814 NNDL-ICP6-700762893.ipynb

colab.research.google.com/drive/1cfoUuUKvBiswUyJfnkQOEKNJKQrMM2#scrollTo=9rOO3Y6L6tfa

Gmail YouTube Udem UCM Single Sign-On ChatGPT Netflix Home - New MyCe... Blackboard Learn SHOPPING BAG Your Shopping Bag... Deep Learning with...

NNDL-ICP6-700762893.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files + Code + Text

sample\_data Sentiment.csv text\_classification...

Requirement already satisfied: wrapt<1.15,>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.14.1)

Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.37.1)

Requirement already satisfied: grpcio<2.0,>=1.24.3 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.64.1)

Requirement already satisfied: tensorflow-estimator<2.16,>=2.15 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.15.2)

Requirement already satisfied: tensorflow-estimator<2.16,>=2.15.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.15.0)

Requirement already satisfied: keras<2.16,>=2.15.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.15.0)

Requirement already satisfied: wheel<1.0,>=0.23.0 in /usr/local/lib/python3.10/dist-packages (from astunparse>=1.6.0->tensorflow) (0.43.0)

Requirement already satisfied: google-auth<3,>=1.6.3 in /usr/local/lib/python3.10/dist-packages (from tensorflowboard<2.16,>=2.15->tensorflow) (2.27.0)

Requirement already satisfied: google-auth-oauthlib<2,>=0.5 in /usr/local/lib/python3.10/dist-packages (from tensorflowboard<2.16,>=2.15->tensorflow) (1.2.1)

Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.10/dist-packages (from tensorflowboard<2.16,>=2.15->tensorflow) (3.6)

Requirement already satisfied: requests<3,>=2.21.0 in /usr/local/lib/python3.10/dist-packages (from tensorflowboard<2.16,>=2.15->tensorflow) (2.31.0)

Requirement already satisfied: tensorflow-data-server<0.8,>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from tensorflowboard<2.16,>=2.15->tensorflow) (0.7.2)

Requirement already satisfied: Werkzeug>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from tensorflowboard<2.16,>=2.15->tensorflow) (3.0.3)

Requirement already satisfied: cachetools<6.0,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorflowboard<2.16,>=2.15->tensorflow) (5.4.0)

Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorflowboard<2.16,>=2.15->tensorflow) (0.4.0)

Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorflowboard<2.16,>=2.15->tensorflow) (4.9)

Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from google-auth-oauthlib<2,>=0.5->tensorflowboard<2.16,>=2.15->tensorflow) (1.3.1)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorflowboard<2.16,>=2.15->tensorflow) (3.3.2)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorflowboard<2.16,>=2.15->tensorflow) (3.7)

Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorflowboard<2.16,>=2.15->tensorflow) (2.0.7)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorflowboard<2.16,>=2.15->tensorflow) (2024.7.4)

Requirement already satisfied: MarkupSafe>=2.1.1 in /usr/local/lib/python3.10/dist-packages (from Werkzeug>=1.0.1->tensorflowboard<2.16,>=2.15->tensorflow) (2.1.5)

Requirement already satisfied: pyasn1<0.7.0,>=0.4.6 in /usr/local/lib/python3.10/dist-packages (from pyasn1-modules>=0.2.1->google-auth<3,>=1.6.3->tensorflowboard<2.16,>=2.15->tensorflow) (0.4.0)

Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.10/dist-packages (from requests-oauthlib>=0.7.0->google-auth-oauthlib<2,>=0.5->tensorflowboard<2.16,>=2.15->tensorflow) (3.2.2)

291/291 - 48s - loss: 0.8276 - accuracy: 0.6428 - 48s/epoch - 164ms/step

144/144 - 4s - loss: 0.7797 - accuracy: 0.6562 - 4s/epoch - 26ms/step

0.779680738639832

0.6561817526817322

['loss', 'accuracy']

Connected to Python 3 Google Compute Engine backend

30°C Mostly sunny