

## Project: Neural Network SMS Text Classifier

Objective:

- create a machine learning model that will classify SMS messages as either 'ham' or 'spam'.
- A 'ham' message is a normal message sent by a friend.
- A 'spam' message is an advertisement or a message sent by a company.

```
In [ ]: import numpy as np
import pandas as pd
from pandas import read_csv
import matplotlib.pyplot as plt
import tensorflow as tf
from tensorflow.keras.preprocessing.sequence import pad_sequences

print(tf.__version__)
```

2023-07-03 22:18:07.154790: I tensorflow/core/platform/cpu\_feature\_guard.cc:182] This TensorFlow binary is optimized to use available CPU instructions in performance-critical operations.  
To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.

## Load dataset

```
In [ ]: # get data files
wget https://cdn.freecodecamp.org/project-data/sms/train-data.tsv
wget https://cdn.freecodecamp.org/project-data/sms/valid-data.tsv

train_file_path = "train-data.tsv"
test_file_path = "valid-data.tsv"

--2023-07-03 22:18:11-- https://cdn.freecodecamp.org/project-data/sms/train-data.tsv
Resolving cdn.freecodecamp.org (cdn.freecodecamp.org)... 104.26.2.33, 104.26.3.33, 172.67.78.149
Connecting to cdn.freecodecamp.org (cdn.freecodecamp.org)|104.26.2.33|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 358233 (358k) [text/tab-separated-values]
Saving to: "train-data.tsv"

train-data.tsv      100%[=====] 349,84K  --.-KB/s  in 0.1s

2023-07-03 22:18:11 (2,78 MB/s) - "train-data.tsv" saved [358233/358233]

--2023-07-03 22:18:12-- https://cdn.freecodecamp.org/project-data/sms/valid-data.tsv
Resolving cdn.freecodecamp.org (cdn.freecodecamp.org)... 104.26.2.33, 104.26.3.33, 172.67.78.149
Connecting to cdn.freecodecamp.org (cdn.freecodecamp.org)|104.26.2.33|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 118774 (118k) [text/tab-separated-values]
Saving to: "valid-data.tsv"

valid-data.tsv      100%[=====] 115,99K  --.-KB/s  in 0.06s

2023-07-03 22:18:12 (1,89 MB/s) - "valid-data.tsv" saved [118774/118774]
```

```
In [ ]: print(train_file_path)
print(test_file_path)

train-data.tsv
valid-data.tsv

In [ ]: df_train = pd.read_csv(train_file_path, sep='\t', header=None, names=['y', 'x'])
df_train.head()
```

```
Out[ ]:      y      x
0  ham  ahhhh...just woken up!had a bad dream about u...
1  ham              you can never do nothing
2  ham  now u sound like manky scouse boy steve!kef...
3  ham  mum say we want to go then go...then she can s...
4  ham  never y lei...i v lazy...got wat? dat day u...
```

```
In [ ]: df_test = pd.read_csv(test_file_path, sep='\t', header=None, names=['y', 'x'])
df_test.head()
```

```
Out[ ]:      y      x
0  ham  iam in hospital da..i will return home in e...
1  ham  not much,just some texin'.how bout you?
2  ham  i probably won't eat at all today.i think i'm...
3  ham  don't give a flying monkeys wot they think and...
4  ham              who are you seeing?
```

```
In [ ]: print(len(df_train))
print(len(df_test))

4179
1392
```

## Handle categorical values

```
In [ ]: y_train = df_train['y'].astype('category').cat.codes
y_test = df_test['y'].astype('category').cat.codes
y_train[:5]
```

```
Out[ ]: 0  0
1  0
2  0
3  0
4  0
dtype: int8
```

```
In [ ]: bar = df_train['y'].value_counts()
```

```
plt.bar(bar.index, bar)
plt.xlabel('Label')
plt.title('Number of ham and spam messages')
```

```
Out[ ]: Text(0.5, 1.0, 'Number of ham and spam messages')
```



## Text preprocessing

```
In [ ]: import nltk
nltk.download('stopwords') # download stopwords
nltk.download('wordnet') # download vocab for Lemmatizer
```

```
[nltk_data] Downloading package stopwords to
[nltk_data] /Users/lenara/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to /Users/lenara/nltk_data...
[nltk_data] Package wordnet is already up-to-date!
```

```
Out[ ]: True
```

```
In [ ]: import re
from nltk.stem import WordNetLemmatizer
from nltk.corpus import stopwords
```

```
In [ ]: stopwords_eng = set(stopwords.words('english'))
len(stopwords_eng)
```

```
Out[ ]: 179
```

```
In [ ]: lemmatizer = WordNetLemmatizer()

def clean_txt(txt):
    txt = re.sub('[^\s\w]+', ' ', txt)
    txt = ''.join([lemmatizer.lemmatize(word) for word in txt.split()
                    if not word in stopwords_eng])
    txt = txt.lower()
    return txt

X_train = df_train['x'].apply(lambda x: clean_txt(x))
X_train[:5]
```

```
Out[ ]: 0  ahhhh woken bad dream u tho dont like u right...
1  i never nothing
2  u sound like manky scouse boy steve like trave...
3  mum say wan go go shun blun watch da glass exh...
4  never lei v lacy got wat dat day u send da url...
Name: x, dtype: object
```

## Vectorize

```
In [ ]: from tensorflow.keras.preprocessing.text import Tokenizer
from keras.preprocessing import sequence
```

```
In [ ]: # Keep top 1000 frequently occurring words
max_words = 1000

# Cut off the words after seeing 500 words in each document
max_len = 500
```

```
In [ ]: t = Tokenizer(num_words=max_words)
t.fit_on_texts(X_train)
```

```
In [ ]: # Transform each text to a sequence of integers
sequences = t.texts_to_sequences(X_train)
sequences[:5]
```

```
Out[ ]: [[189, 227, 1, 587, 42, 15, 1, 90, 359, 13, 103, 54, 228, 86],
 [195, 252],
 [1, 310, 15, 219, 15, 43, 311, 37, 386, 1, 6, 338, 422],
 [477, 58, 188, 8, 8, 243, 43],
 [195, 478, 187, 48, 77, 212, 12, 28, 22, 43, 124, 78, 24]]
```

```
In [ ]: # Make all rows of equal length
sequences_matrix = pad_sequences(sequences, maxlen=max_len)
sequences_matrix[:5]
```

```
Out[ ]: array([[ 0,  0,  0, ..., 54, 228, 86],
 [ 0,  0,  0, ...,  0, 195, 252],
 [ 0,  0,  0, ...,  0, 338, 422],
 [ 0,  0,  0, ...,  8, 243, 43],
 [ 0,  0,  0, ..., 124, 78, 24]], dtype=int32)
```

## Build model

```
In [ ]: i = tf.keras.layers.Input(shape=(max_len,))
x = tf.keras.layers.Embedding(max_words, 50, input_length=max_len)(i)
x = tf.keras.layers.LSTM(64)(x)

x = tf.keras.layers.Dense(256, activation='relu')(x)
x = tf.keras.layers.Dropout(0.5)(x)
x = tf.keras.layers.Dense(1, activation='relu')(x)

model = tf.keras.models.Model(inputs=i, outputs=x)
model.compile(
    loss='binary_crossentropy',
    optimizer='Adam',
    metrics=['accuracy']
)
model.summary()
```

Model: "model"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 500)]	0
embedding_1 (Embedding)	(None, 500, 50)	50000
lstm_1 (LSTM)	(None, 64)	29440
dense_1 (Dense)	(None, 256)	16640
dropout_1 (Dropout)	(None, 256)	0
dense_1_1 (Dense)	(None, 1)	257

Total params: 96,337  
Trainable params: 96,337  
Non-trainable params: 0

2023-07-03 22:18:18.829142: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_2\_grad/concat/split\_2/split\_dim' with dtype int32  
[[[node gradients/split\_2\_grad/concat/split\_2/split\_dim]]]

2023-07-03 22:18:18.834851: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:18.835877: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

```
In [ ]: m = model.fit(sequences_matrix, y_train,
    batch_size=128, epochs=10,
    validation_split=0.2,
    callbacks=[tf.keras.callbacks.EarlyStopping(
        monitor='val_loss', min_delta=0.0001)])
```

Epoch 1/10  
2023-07-03 22:18:18.829197: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_2\_grad/concat/split\_2/split\_dim' with dtype int32  
[[[node gradients/split\_2\_grad/concat/split\_2/split\_dim]]]

2023-07-03 22:18:18.831540: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.786239: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.786808: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.714429: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.323580: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_2\_grad/concat/split\_2/split\_dim' with dtype int32  
[[[node gradients/split\_2\_grad/concat/split\_2/split\_dim]]]

2023-07-03 22:18:19.324382: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]

2023-07-03 22:18:19.326431: I tensorflow/core/common\_runtime/executor.cc:1197] [device:CPU:0] (DEBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID\_ARGUMENT: You must feed a value for placeholder tensor 'gradients/split\_1\_grad/concat/split\_1/split\_dim' with dtype int32  
[[[node gradients/split\_1\_grad/concat/split\_1/split\_dim]]]