

Scrip3

February 29, 2024

SENTIMENT ANALYSIS with Transformers

0.1 Project [GitHub](#)

0.2 [Link](#) To Our Model Weights After Training and Dataset Google Drive

```
[1]: #@title Main Imports for Functionality

import os
import re

from typing import Union, List
from tqdm import tqdm

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split

import torch

import nltk
nltk.download("stopwords")
from nltk.corpus import stopwords

import pandas as pd
from wordcloud import WordCloud
import seaborn as sns

import string
from collections import Counter, defaultdict

from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer

import plotly.express as px
from plotly.subplots import make_subplots
import plotly.graph_objects as go
```

```

from plotly.offline import plot

import matplotlib.gridspec as gridspec
from matplotlib.ticker import MaxNLocator
import matplotlib.patches as mpatches
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')

import pandas as pd
import numpy as np
import os
import random
from pathlib import Path
import json
import torch
from tqdm.notebook import tqdm

from transformers import BertTokenizer
from torch.utils.data import TensorDataset

from transformers import BertForSequenceClassification

%matplotlib inline

```

[nltk_data] Downloading package stopwords to /root/nltk_data...

[nltk_data] Unzipping corpora/stopwords.zip.

```

[2]: from google.colab import drive
drive.mount('/content/drive')

```

Mounted at /content/drive

```

[3]: os.chdir("/content/drive/MyDrive/NLP/")

```

```

[ ]: #@title Acquire and convert sentiment to integers
data_all = pd.read_excel('/content/drive/MyDrive/NLP/IMDB Dataset.xlsx')

# # splitting dataframe by row index
# set_validate = data_all.iloc[:5000,:]
# set_training = data_all.iloc[5000:,:]
# print("Shape of new dataframes - {} , {}".format(set_validate.shape,
# ↪set_training.shape))

# prepare Training set and view first 3 entries
sentiment_list = list(data_all["sentiment"])
sentiment_binary = []

```

```

for item in sentiement_list:
    if item == "positive":
        sentiement_binary.append(1)
    else:
        sentiement_binary.append(0)

sentiement_list = list(data_all["sentiment"])

data_all["label"] = sentiement_binary
data_all.drop(["sentiment"], inplace=True, axis=1)

print(data_all.sample(3))
print(" ")

# # prepare Validation set and view first 3 entries
# set_validate.drop(["sentiment"], inplace=True, axis=1)

# print(set_validate.sample(3))
# print(" ")

```

	review	label
17143	The Hollow is a wonderful murder mystery that ...	1
29274	One of the best records of Israel's response t...	1
9292	This film revolves as much around Japanese cul...	1

2. EDA

```

[ ]: import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
stopWords_nltk = set(stopwords.words('english'))

```

```

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!

```

```

[ ]: # Check GPU availability Stuff

if torch.cuda.is_available():
    device = torch.device("cuda")
    print(f'There are {torch.cuda.device_count()} GPU(s) available.')
    print('Device name:', torch.cuda.get_device_name(0))

else:
    print('No GPU available, using the CPU instead.')
    device = torch.device("cpu")

```

There are 1 GPU(s) available.
Device name: Tesla T4

```
[ ]: # Text preprocessing

def text_preprocessing(s):
    """
    - Lowercase the sentence
    - Change "'t" to "not"
    - Remove "@name"
    - Isolate and remove punctuations except "?"
    - Remove other special characters
    - Remove stop words except "not" and "can"
    - Remove trailing whitespace
    """
    s = s.lower()
    # Change 't to 'not'
    s = re.sub(r"'t", " not", s)
    # Remove @name
    s = re.sub(r'(@.*?)[\s]', ' ', s)
    # Isolate and remove punctuations except '?'
    s = re.sub(r'([\\".\(\)\!\\?\\\/\,])', r' ', s)
    s = re.sub(r'[\w\s\?]', ' ', s)
    # Remove some special characters
    s = re.sub(r'([\;\:\|\•<\n]<br /><br />)', ' ', s)
    # Remove stopwords except 'not' and 'can'
    s = " ".join([word for word in s.split()
                   if word not in stopwords.words('english')
                   or word in ['not', 'can']])
    # Remove trailing whitespace
    s = re.sub(r'\s+', ' ', s).strip()

    return s
```

```
[ ]: data_all = pd.read_excel('/content/drive/MyDrive/NLP/data_all.xlsx')
```

```
[ ]: data_all.head()
```

```
[ ]:   label                                review2
0      1  one reviewers mentioned watching 1 oz episode ...
1      1  wonderful little production br br filming tech...
2      1  thought wonderful way spend time hot summer we...
3      0  basically family little boy jake thinks zombie...
4      1  petter mattei love time money visually stunnin...
```

```
[ ]: # basic info
data_all.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  -
0   label       50000 non-null  int64
1   review2     50000 non-null  object
dtypes: int64(1), object(1)
memory usage: 781.4+ KB

```

```

[ ]: # Preprocess text
data_all['review2'] = ([text_preprocessing(text) for text in
    ↪data_all['review']])

```

```

[ ]: data_all.drop(["review"], inplace=True, axis=1)

```

```

[ ]: data_all.to_excel('/content/drive/MyDrive/NLP/data_all.xlsx', index=False)

```

Visualizations

0.3 Word Cloud

```

[ ]: def show_wordcloud(data, title = None):
    wordcloud = WordCloud(
        background_color='black',
        max_words=100,
        max_font_size=40,
        scale=1,
        random_state=1
    ).generate(" ".join(data))

    fig = plt.figure(1, figsize=(15, 15))
    plt.axis('off')
    if title:
        fig.suptitle(title, fontsize=20)
        fig.subplots_adjust(top=2.3)

    plt.imshow(wordcloud)
    plt.show()

```

```

[ ]: show_wordcloud(data_all["review2"])

```


0.5 BERT tokenizer

```
[ ]: from transformers import BertTokenizer
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased',
                                         do_lower_case=True)

tokenizer_config.json: 0%|          | 0.00/48.0 [00:00<?, ?B/s]
vocab.txt: 0%|          | 0.00/232k [00:00<?, ?B/s]
tokenizer.json: 0%|          | 0.00/466k [00:00<?, ?B/s]
config.json: 0%|          | 0.00/570 [00:00<?, ?B/s]

[ ]: # data tokenize with bert tokenizer
data_all["Token_length"] = data_all["review2"].apply(lambda x: len(tokenizer(x,
↪add_special_tokens=False)["input_ids"])))
```

Token indices sequence length is longer than the specified maximum sequence length for this model (582 > 512). Running this sequence through the model will result in indexing errors

```
[ ]: fig = px.histogram(data_all, x="Token_length", nbins=20,
↪color_discrete_sequence=px.colors.cmocean.algae, barmode='group',
↪histnorm="percent")
fig.show()
```

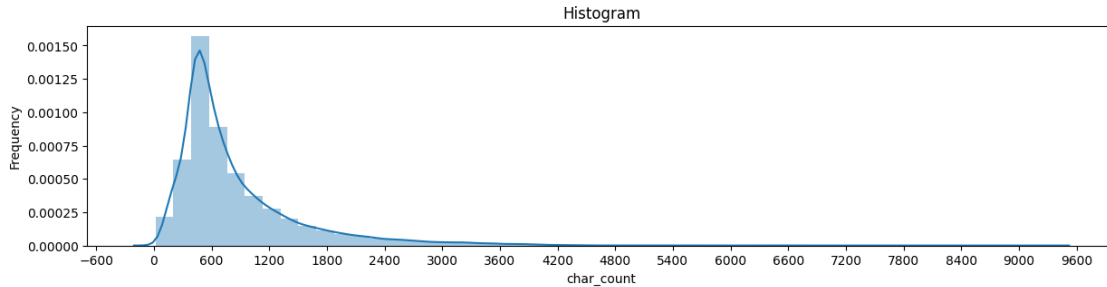
0.6 Characters Count in the Data

Let's look at the frequency of the number of characters. It will give us information about the overall size of our data

```
[ ]: # valvulate char count for each review
data_all['char_count'] = data_all['review2'].apply(lambda x: len(str(x)))

def plot_dist3(df, feature):
    fig = plt.figure(constrained_layout=True, figsize=(18, 8))
    grid = gridspec.GridSpec(ncols=3, nrows=3, figure=fig)
    ax1 = fig.add_subplot(grid[0, :2])
    ax1.set_title('Histogram')
    sns.distplot(df.loc[:, feature],
                 hist=True,
                 kde=True,
                 ax=ax1,
                 )
    ax1.set_ylabel('Frequency')
    ax1.xaxis.set_major_locator(MaxNLocator(nbins=20))

[ ]: plot_dist3(data_all, 'char_count')
```



0.7 Reviews Lengths

```
[ ]: # Creating a new feature for the visualization.
data_all['Character Count'] = data_all['review2'].apply(lambda x: len(str(x)))

def plot_dist3(df, feature, title):
    # Creating a customized chart. and giving in figsize and everything.
    fig = plt.figure(constrained_layout=True, figsize=(24, 12))
    # Creating a grid of 3 cols and 3 rows.
    grid = gridspec.GridSpec(ncols=3, nrows=3, figure=fig)

    # Customizing the histogram grid.
    ax1 = fig.add_subplot(grid[0, :2])
    # Set the title.
    ax1.set_title('Histogram')
    # plot the histogram.
    sns.distplot(df.loc[:, feature],
                  hist=True,
                  kde=True,
                  ax=ax1,
                  color='#e74c3c')
    ax1.set(ylabel='Frequency')
    ax1.xaxis.set_major_locator(MaxNLocator(nbins=20))

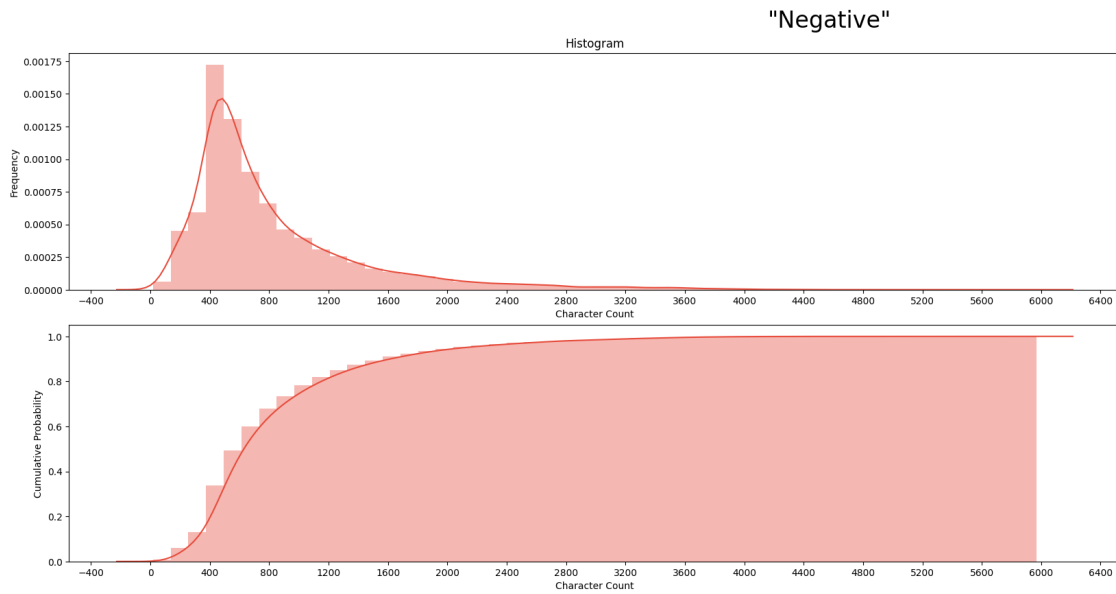
    ax2 = fig.add_subplot(grid[1, :2])

    sns.distplot(df.loc[:, feature],
                  ax=ax2,
                  kde_kws={'cumulative': True},
                  hist_kws={'cumulative': True},
                  color='#e74c3c')
    ax2.xaxis.set_major_locator(MaxNLocator(nbins=20))
    ax2.set(ylabel='Cumulative Probability')
```

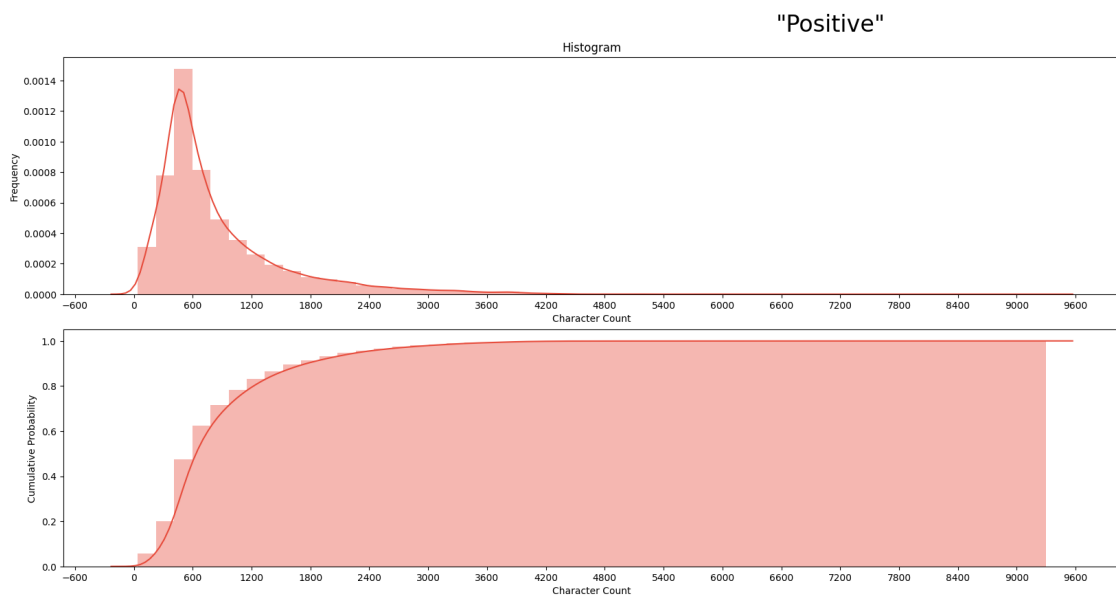


```
plt.suptitle(f'{title}', fontsize=24)
```

```
[ ]: plot_dist3(data_all[data_all['label'] == 0], 'Character Count',  
                '"Negative"')
```



```
[ ]: plot_dist3(data_all[data_all['label'] == 1], 'Character Count',  
                '"Positive"')
```



0.8 Word Counts

```
[ ]: def plot_word_number_histogram(textno, textye):  
  
    """A function for comparing word counts"""  
  
    fig, axes = plt.subplots(ncols=1, nrows=2, figsize=(18, 12), sharey=True)  
    sns.distplot(textno.str.split().map(lambda x: len(x)), ax=axes[0],  
↳color='#e74c3c')  
    sns.distplot(textye.str.split().map(lambda x: len(x)), ax=axes[1],  
↳color='#e74c3c')  
  
    axes[0].set_xlabel('Word Count')  
    axes[0].set_ylabel('Frequency')  
    axes[0].set_title('negative')  
    axes[1].set_xlabel('Word Count')  
    axes[1].set_title('positive')  
  
    fig.suptitle('Words Per Review', fontsize=24, va='baseline')  
  
    fig.tight_layout()
```

```
[ ]: plot_word_number_histogram(data_all[data_all['label'] == 0]['review2'],  
                               data_all[data_all['label'] == 1]['review2'] )
```



Most Common Words

```
[ ]: texts = data_all['review2']
new = texts.str.split()
new = new.values.tolist()
corpus = [word for i in new for word in i]
counter = Counter(corpus)
most = counter.most_common()
x, y = [], []
for word, count in most[:30]:
    if word not in stopWords_nltk:
        x.append(word)
        y.append(count)

fig = go.Figure(go.Bar(
    x=y,
    y=x,
    orientation='h', marker=dict(
        color='rgba(50, 171, 96, 0.6)',
        line=dict(
            color='rgba(50, 171, 96, 1.0)',
            width=1),
    ),
    name='Most common Word',))

fig.update_layout( title={
    'text': "Most Common Words",
    'y':0.9,
    'x':0.5,
    'xanchor': 'center',
    'yanchor': 'top'}, font=dict(
    family="Courier New, monospace",
    size=18,
    color="RebeccaPurple"
))

fig.show()
```

0.9 Most Common ngrams

```
[ ]: fig = make_subplots(rows=1, cols=2)
title_ = ["negative", "positive"]

for i in range(2):
    texts = data_all['review2']
```

```

new = texts.str.split()
new = new.values.tolist()
corpus = [word for i in new for word in i]
counter = Counter(corpus)
most = counter.most_common()
x, y = [], []

for word, count in most[:30]:
    if word not in stopWords_nltk:
        x.append(word)
        y.append(count)

fig.add_trace(go.Bar(
    x=y,
    y=x,
    orientation='h', type="bar",
    name=title_[i], marker=dict(color=colors[i])), 1, i+1)

fig.update_layout(
    autosize=False,
    width=2000,
    height=600, title=dict(
        text='<b>Most Common ngrams per Classes</b>',
        x=0.5,
        y=0.95,
        font=dict(
            size=24,
            color="RebeccaPurple"
        )
    ),)

fig.show()

```

```
[ ]: data_all.head()
```

```
[ ]:
```

	label	review2	char_count	\
0	1	one reviewers mentioned watching 1 oz episode ...	1151	
1	1	wonderful little production br br filming tech...	686	
2	1	thought wonderful way spend time hot summer we...	594	
3	0	basically family little boy jake thinks zombie...	456	
4	1	petter mattei love time money visually stunnin...	880	

	Token_length
0	193
1	107
2	98

```
3         80
4         153
```

```
[ ]: def _get_top_ngram(corpus, n=None):
    #getting top ngrams
    vec = CountVectorizer(ngram_range=(n, n),
                          max_df=0.9,
                          ).fit(corpus)
    bag_of_words = vec.transform(corpus)
    sum_words = bag_of_words.sum(axis=0)
    words_freq = [(word, sum_words[0, idx])
                   for word, idx in vec.vocabulary_.items()]
    words_freq = sorted(words_freq, key=lambda x: x[1], reverse=True)
    return words_freq[:15]
```

```
[ ]: # unigram
fig = make_subplots(rows=1, cols=2)

title_ = ["negative", "positive"]

for i in range(2):
    texts = data_all[data_all["label"] == i]['review2']

    new = texts.str.split()
    new = new.values.tolist()
    corpus = [word for i in new for word in i]
    top_n_bigrams = _get_top_ngram(texts, 2)[:15]
    x, y = map(list, zip(*top_n_bigrams))

    fig.add_trace(go.Bar(
        x=y,
        y=x,
        orientation='h', type="bar",
        name=title_[i], marker=dict(color=colors[i])), 1, i+1)

fig.update_layout(
    autosize=False,
    width=2000,
    height=600, title=dict(
        text='<b>Most Common unigrams per Classes</b>',
        x=0.5,
        y=0.95,
        font=dict(
            family="Courier New, monospace",
            size=24,
```

```

        color="RebeccaPurple"
    )
))
fig.show()

```

```

[ ]: #trigram

fig = make_subplots(rows=1, cols=2)
title_ = ["negative", "positive"]

for i in range(2):
    texts = data_all[data_all["label"] == i]['review2']

    new = texts.str.split()
    new = new.values.tolist()
    corpus = [word for i in new for word in i]

    top_n_bigrams = _get_top_ngram(texts, 3)[:15]
    x, y = map(list, zip(*top_n_bigrams))

    fig.add_trace(go.Bar(
        x=y,
        y=x,
        orientation='h', type="bar",
        name=title_[i], marker=dict(color=colors[i])), 1, i+1),

fig.update_layout(
    autosize=False,
    width=2000,
    height=600, title=dict(
        text='<b>Most Common trigrams per Classes</b>',
        x=0.5,
        y=0.95,
        font=dict(
            family="Courier New, monospace",
            size=24,
            color="RebeccaPurple"
        )
    ))

fig.show()

```

Transformer Model

BERT Train

```

[ ]: class Config():
    seed_val = 17

```

```

device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
epochs = 1
batch_size = 6
seq_length = 512
lr = 2e-5
eps = 1e-8
pretrained_model = 'bert-base-uncased'
test_size=0.15
random_state=42
add_special_tokens=True
return_attention_mask=True
pad_to_max_length=True
do_lower_case=False
return_tensors='pt'

config = Config()

```

0.10 Params we will be saving

```

[ ]: params = {"seed_val": config.seed_val,
               "device":str(config.device),
               "epochs":config.epochs,
               "batch_size":config.batch_size,
               "seq_length":config.seq_length,
               "lr":config.lr,
               "eps":config.eps,
               "pretrained_model": config.pretrained_model,
               "test_size":config.test_size,
               "random_state":config.random_state,
               "add_special_tokens":config.add_special_tokens,
               "return_attention_mask":config.return_attention_mask,
               "pad_to_max_length":config.pad_to_max_length,
               "do_lower_case":config.do_lower_case,
               "return_tensors":config.return_tensors,
               }

```

```

[ ]: import random

device = config.device

random.seed(config.seed_val)
np.random.seed(config.seed_val)
torch.manual_seed(config.seed_val)
torch.cuda.manual_seed_all(config.seed_val)

```

Train and Validation Split

```
[ ]: from sklearn.model_selection import train_test_split
```

```
train_df, val= train_test_split(data_all,  
                                test_size=0.10,  
                                random_state=config.random_state,  
                                stratify=data_all.label.values)
```

```
[ ]: train_df.head()
```

```
[ ]:      label                                review2  Token_length \  
14700      0  since little mermaid one favorite disney movie...      211  
10733      0  based novel michael chabon mysteries pittsburg...      114  
30659      0  makes third errol morris movie seen increasing...      232  
10987      0  sitting big wing chair huge book lap one bela ...      365  
33167      1  kid movie great adult mother enjoyed watching ...      162
```

```
      Character Count  
14700              1334  
10733              714  
30659             1465  
10987             2112  
33167             1034
```

```
[ ]: val.head()
```

```
[ ]:      label                                review2  Token_length \  
30859      1  not kidding summary vote video distributors pa...      268  
7217       0  kept waiting film move inspire shock sadden wa...      61  
34889      0  movie 90 minute ramones concert brief periods ...      55  
30806      1  every scene put together perfectly movie wonde...      119  
36408      0  uma travolta good together unfortunately left ...      96
```

```
      Character Count  
30859             1595  
7217              377  
34889             314  
30806             741  
36408             543
```

```
[ ]: train, test = train_test_split(train_df,  
                                    test_size=0.10,  
                                    random_state=42,  
                                    stratify=train_df.label.values)
```

```
[ ]: # count of unique label control  
print(len(train['label'].unique()))  
print(train.shape)
```



```
2
(40500, 4)
```

```
[ ]: # count of unique label control
print(len(test['label'].unique()))
print(test.shape)
```

```
2
(4500, 4)
```

```
[ ]: # count of unique label control
print(len(val['label'].unique()))
print(val.shape)
```

```
2
(5000, 4)
```

BertTokenizer

```
[ ]: # create tokenizer
tokenizer = BertTokenizer.from_pretrained(config.pretrained_model,
                                         do_lower_case=config.do_lower_case)
```

```
[ ]: encoded_data_train = tokenizer.batch_encode_plus(
    train.review2.values,
    add_special_tokens=config.add_special_tokens,
    return_attention_mask=config.return_attention_mask,
    pad_to_max_length=config.pad_to_max_length,
    max_length=config.seq_length,
    return_tensors=config.return_tensors
)
encoded_data_val = tokenizer.batch_encode_plus(
    val.review2.values,
    add_special_tokens=config.add_special_tokens,
    return_attention_mask=config.return_attention_mask,
    pad_to_max_length=config.pad_to_max_length,
    max_length=config.seq_length,
    return_tensors=config.return_tensors
)
```

Truncation was not explicitly activated but `max_length` is provided a specific value, please use `truncation=True` to explicitly truncate examples to max length. Defaulting to 'longest_first' truncation strategy. If you encode pairs of sequences (GLUE-style) with the tokenizer you can select this strategy more precisely by providing a specific strategy to `truncation`.

```
[ ]: input_ids_train = encoded_data_train['input_ids']
attention_masks_train = encoded_data_train['attention_mask']
labels_train = torch.tensor(train.label.values)
```

```
input_ids_val = encoded_data_val['input_ids']
attention_masks_val = encoded_data_val['attention_mask']
labels_val = torch.tensor(val.label.values)
```

```
[ ]: dataset_train = TensorDataset(input_ids_train, attention_masks_train,
    ↪ labels_train)
dataset_val = TensorDataset(input_ids_val, attention_masks_val, labels_val)
```

Creating the Model

```
[ ]: model = BertForSequenceClassification.from_pretrained(config.pretrained_model,
    num_labels=2,
    output_attentions=False,
    ↪ output_hidden_states=False)
```

```
model.safetensors: 0%|          | 0.00/440M [00:00<?, ?B/s]
```

Some weights of BertForSequenceClassification were not initialized from the model checkpoint at bert-base-uncased and are newly initialized:

```
['classifier.bias', 'classifier.weight']
```

You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.

Data Loaders

```
[ ]: from torch.utils.data import DataLoader, RandomSampler, SequentialSampler

dataloader_train = DataLoader(dataset_train,
    sampler=RandomSampler(dataset_train),
    batch_size=config.batch_size)

dataloader_validation = DataLoader(dataset_val,
    sampler=SequentialSampler(dataset_val),
    batch_size=config.batch_size)
```

Optimizer & Scheduler

```
[ ]: from transformers import AdamW, get_linear_schedule_with_warmup

optimizer = AdamW(model.parameters(),
    lr=config.lr,
    eps=config.eps)

scheduler = get_linear_schedule_with_warmup(optimizer,
    num_warmup_steps=0,
```

```

↳ num_training_steps=len(dataloader_train)*config.epochs

```

Performance Metrics

We will use f1 score as performance metrics.

```

[ ]: from sklearn.metrics import f1_score

def f1_score_func(preds, labels):
    preds_flat = np.argmax(preds, axis=1).flatten()
    labels_flat = labels.flatten()
    return f1_score(labels_flat, preds_flat, average='weighted')

def accuracy_per_class(preds, labels, label_dict):
    label_dict_inverse = {v: k for k, v in label_dict.items()}

    preds_flat = np.argmax(preds, axis=1).flatten()
    labels_flat = labels.flatten()

    for label in np.unique(labels_flat):
        y_preds = preds_flat[labels_flat==label]
        y_true = labels_flat[labels_flat==label]
        print(f'Class: {label_dict_inverse[label]}')
        print(f'Accuracy: {len(y_preds[y_preds==label])}/{len(y_true)}\n')

```

Training Loop

```

[ ]: def evaluate(dataloader_val):

    model.eval()

    loss_val_total = 0
    predictions, true_vals = [], []

    for batch in dataloader_val:

        batch = tuple(b.to(config.device) for b in batch)

        inputs = {'input_ids':      batch[0],
                  'attention_mask': batch[1],
                  'labels':         batch[2],
                  }

        with torch.no_grad():
            outputs = model(**inputs)

        loss = outputs[0]

```

```

logits = outputs[1]
loss_val_total += loss.item()

logits = logits.detach().cpu().numpy()
label_ids = inputs['labels'].cpu().numpy()
predictions.append(logits)
true_vals.append(label_ids)

# calculate avareage val loss
loss_val_avg = loss_val_total/len(dataloader_val)

predictions = np.concatenate(predictions, axis=0)
true_vals = np.concatenate(true_vals, axis=0)

return loss_val_avg, predictions, true_vals

```

```
[ ]: config.device
```

```
[ ]: device(type='cuda', index=0)
```

```

[ ]: model.to(config.device)

for epoch in tqdm(range(1, 2)):

    model.train()

    loss_train_total = 0

    progress_bar = tqdm(dataloader_train, desc='Epoch {:1d}'.format(epoch),
↳leave=False, disable=False)

    for batch in progress_bar:

        model.zero_grad()

        batch = tuple(b.to(config.device) for b in batch)

        inputs = {'input_ids':      batch[0],
                  'attention_mask': batch[1],
                  'labels':         batch[2],
                  }

        outputs = model(**inputs)

        loss = outputs[0]
        loss_train_total += loss.item()

```

```

        loss.backward()

        torch.nn.utils.clip_grad_norm_(model.parameters(), 1.0)

        optimizer.step()
        scheduler.step()

        progress_bar.set_postfix({'training_loss': '{:.3f}'.format(loss.item()/
↪len(batch))})

        torch.save(model.state_dict(), f'/content/drive/MyDrive/NLP/
↪BERT_epoch_{epoch}.model')

        tqdm.write(f'\nEpoch {epoch}')

        loss_train_avg = loss_train_total/len(dataloader_train)
        tqdm.write(f'Training loss: {loss_train_avg}')

        val_loss, predictions, true_vals = evaluate(dataloader_validation)
        val_f1 = f1_score_func(predictions, true_vals)
        tqdm.write(f'Validation loss: {val_loss}')

        tqdm.write(f'F1 Score (Weighted): {val_f1}');
# save model params and other configs
        with Path('/content/drive/MyDrive/NLP/params.json').open("w") as f:
            json.dump(params, f, ensure_ascii=False, indent=4)

```

```

0%|          | 0/1 [00:00<?, ?it/s]
Epoch 1:  0%|          | 0/6750 [00:00<?, ?it/s]

```

```

-----
KeyboardInterrupt                                Traceback (most recent call last)
<ipython-input-35-588152d5240f> in <cell line: 3>()
    29         torch.nn.utils.clip_grad_norm_(model.parameters(), 1.0)
    30
----> 31         optimizer.step()
    32         scheduler.step()
    33

/usr/local/lib/python3.10/dist-packages/torch/optim/lr_scheduler.py in
↪wrapper(*args, **kwargs)
    66             instance._step_count += 1
    67             wrapped = func.__get__(instance, cls)
----> 68             return wrapped(*args, **kwargs)
    69

```

```

70          # Note that the returned function here is no longer a bound
↪method,

/usr/local/lib/python3.10/dist-packages/torch/optim/optimizer.py in ↪
↪wrapper(*args, **kwargs)
371         )
372
--> 373         out = func(*args, **kwargs)
374         self._optimizer_step_code()
375

/usr/local/lib/python3.10/dist-packages/torch/utils/_contextlib.py in ↪
↪decorate_context(*args, **kwargs)
113     def decorate_context(*args, **kwargs):
114         with ctx_factory():
--> 115             return func(*args, **kwargs)
116
117     return decorate_context

/usr/local/lib/python3.10/dist-packages/transformers/optimization.py in ↪
↪step(self, closure)
484         exp_avg.mul_(beta1).add_(grad, alpha=(1.0 - beta1))
485         exp_avg_sq.mul_(beta2).addcmul_(grad, grad, value=1.0 - ↪
↪beta2)
--> 486         denom = exp_avg_sq.sqrt().add_(group["eps"])
487
488         step_size = group["lr"]

KeyboardInterrupt:

```

Test on validation set

```

[ ]: model.to(config.device)

model.load_state_dict(torch.load('/content/drive/MyDrive/NLP/BERT_epoch_1.
↪model'))
# device = torch.device("cuda")
# # model = TheModelClass(*args, **kwargs)
# model.load_state_dict(torch.load(f'/content/drive/MyDrive/NLP/BERT_epoch_1.
↪model', map_location="cuda:0")) # Choose whatever GPU device number you want

```

[]: <All keys matched successfully>

Evaluation

```

[ ]: pred_final = []

for i, row in tqdm(val.iterrows(), total=val.shape[0]):

```

```

predictions = []

review = row["review2"]
encoded_data_test_single = tokenizer.batch_encode_plus(
    [review],
    add_special_tokens=config.add_special_tokens,
    return_attention_mask=config.return_attention_mask,
    pad_to_max_length=config.pad_to_max_length,
    max_length=config.seq_length,
    return_tensors=config.return_tensors
)
input_ids_test = encoded_data_test_single['input_ids']
attention_masks_test = encoded_data_test_single['attention_mask']

inputs = {'input_ids': input_ids_test.to(device),
          'attention_mask': attention_masks_test.to(device),
          }

with torch.no_grad():
    outputs = model(**inputs)

logits = outputs[0]
logits = logits.detach().cpu().numpy()
predictions.append(logits)
predictions = np.concatenate(predictions, axis=0)
pred_final.append(np.argmax(predictions, axis=1).flatten()[0])

```

```
0%|          | 0/429 [00:00<?, ?it/s]
```

```
[ ]: val["pred"] = pred_final
```

```
[ ]: control = val.pred.values == val.label.values
val["control"] = control
```

```
[ ]: #adding prediction col
val["pred"] = pred_final
control = val.pred.values == val.label.values
val["control"] = control
```

```
[ ]: # filtering predictions
val = val[val.control == False]
```

```
[ ]: val = val[val.control == False]
```

```
[ ]: name2label = {"Negative":0,
                  "Positive":1}
```

```

    }
    label2name = {v: k for k, v in name2label.items()}

    val["pred_name"] = val.pred.apply(lambda x: label2name.get(x))

```

```

[ ]: from sklearn.metrics import confusion_matrix

# We create a confusion matrix to better observe the classes that the model
    ↳confuses.
pred_name_values = val.pred_name.values
label_values = val.pred_name.values
confmat = confusion_matrix(label_values, pred_name_values,
    ↳labels=list(name2label.keys()))

```

```

[ ]: confmat

```

```

[ ]: array([[236,  0],
          [ 0, 144]])

```

```

[ ]: df_confusion_val = pd.crosstab(label_values, pred_name_values)
df_confusion_val

```

```

[ ]: col_0      Negative  Positive
row_0
Negative      236         0
Positive       0        144

```

```

[ ]: df_confusion_val.to_csv("val_df_confusion.csv")

```

```

[ ]: test.head()

```

```

[ ]:
      label      review2  Token_length \
22601     0  one entertaining flick suggest rent buy couple...      70
36824     0  2000 came close king kong adopted daughter wen...     165
48367     1  anatomie certainly one better movies seen not ...      55
15745     0  ravaged wasteland future mankind terrorized cy...     133
22034     1  big gone wind nut disappointed gone wind movie...      89

```

```

      Character Count  pred  control  pred_name
22601              428    1    False  Positive
36824              951    1    False  Positive
48367              338    0    False  Negative
15745              689    1    False  Positive
22034              555    0    False  Negative

```

```

[ ]: encoded_data_test = tokenizer.batch_encode_plus(
      test.review2.values,

```



```

        add_special_tokens=config.add_special_tokens,
        return_attention_mask=config.return_attention_mask,
        pad_to_max_length=config.pad_to_max_length,
        max_length=config.seq_length,
        return_tensors=config.return_tensors
    )

```

```

[ ]: input_ids_test = encoded_data_test['input_ids']
     attention_masks_test = encoded_data_test['attention_mask']
     labels_test = torch.tensor(test.label.values)

```

```

[ ]: model = BertForSequenceClassification.from_pretrained(config.pretrained_model,
                                                         num_labels=2,
                                                         output_attentions=False,
                                                         ↪output_hidden_states=False)

model.to(config.device)

model.load_state_dict(torch.load(f'/content/drive/MyDrive/NLP/BERT_epoch_1.
    ↪model', map_location=torch.device('cpu'))))

_, predictions_test, true_vals_test = evaluate(dataloader_validation)

```

Some weights of BertForSequenceClassification were not initialized from the model checkpoint at bert-base-uncased and are newly initialized:

['classifier.bias', 'classifier.weight']

You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.

```

[ ]: from sklearn.metrics import classification_report

     preds_flat_test = np.argmax(predictions_test, axis=1).flatten()
     print(classification_report(preds_flat_test, true_vals_test))

```

	precision	recall	f1-score	support
0	0.94	0.90	0.92	2604
1	0.90	0.93	0.91	2396
accuracy			0.92	5000
macro avg	0.92	0.92	0.92	5000
weighted avg	0.92	0.92	0.92	5000

```

[ ]: pred_final = []

```

```

for i, row in tqdm(test.iterrows(), total=test.shape[0]):
    predictions = []

    review = row["review2"]
    encoded_data_test_single = tokenizer.batch_encode_plus(
        [review],
        add_special_tokens=config.add_special_tokens,
        return_attention_mask=config.return_attention_mask,
        pad_to_max_length=config.pad_to_max_length,
        max_length=config.seq_length,
        return_tensors=config.return_tensors
    )
    input_ids_test = encoded_data_test_single['input_ids']
    attention_masks_test = encoded_data_test_single['attention_mask']

    inputs = {'input_ids': input_ids_test.to(device),
              'attention_mask': attention_masks_test.to(device),
              }

    with torch.no_grad():
        outputs = model(**inputs)

    logits = outputs[0]
    logits = logits.detach().cpu().numpy()
    predictions.append(logits)
    predictions = np.concatenate(predictions, axis=0)
    pred_final.append(np.argmax(predictions, axis=1).flatten()[0])

```

```
0%|          | 0/375 [00:00<?, ?it/s]
```

```

[ ]: test["pred"] = pred_final
control = test.pred.values == test.label.values
test["control"] = control

```

```

[ ]: test = test[test.control == False]

```

```

[ ]: test["pred_name"] = test.pred.apply(lambda x: label2name.get(x))

```

```

[ ]: from sklearn.metrics import confusion_matrix

# We create a confusion matrix to better observe the classes that the model
↳ confuses.
pred_name_values = test.pred_name.values
label_values = test.pred_name.values
confmat = confusion_matrix(label_values, pred_name_values,
↳ labels=list(name2label.keys()))

```

```
[ ]: confmat
```

```
[ ]: array([[255,  0],  
          [ 0, 120]])
```

```
[ ]: df_confusion_test = pd.crosstab(label_values, pred_name_values)  
df_confusion_test
```

```
[ ]: col_0      Negative  Positive  
row_0  
Negative      255         0  
Positive       0        120
```

```
[ ]: !apt-get install texlive texlive-xetex texlive-latex-extra pandoc  
!pip install py pandoc
```

```
Reading package lists... Done  
Building dependency tree... Done  
Reading state information... Done  
pandoc is already the newest version (2.9.2.1-3ubuntu2).  
pandoc set to manually installed.  
The following additional packages will be installed:  
  dvismgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-  
texgyre  
  fonts-urw-base35 libapache-pom-java libcommons-logging-java libcommons-parent-  
java  
  libfontbox-java libfontenc1 libgs9 libgs9-common libidn12 libijs-0.35  
libjbig2dec0 libkpathsea6  
  libpdfbox-java libptexenc1 libruby3.0 libsynchronet2 libteckit0 libtexlua53  
libtexlua53 libwoff1  
  libzzip-0-13 lmodern poppler-data preview-latex-style rake ruby ruby-net-  
telnet ruby-rubygems  
  ruby-webrick ruby-xmlrpc ruby3.0 rubygems-integration t1utils teckit tex-  
common tex-gyre  
  texlive-base texlive-binaries texlive-fonts-recommended texlive-latex-base  
  texlive-latex-recommended texlive-pictures texlive-plain-generic tipa xfonts-  
encodings  
  xfonts-utils  
Suggested packages:  
  fonts-noto fonts-freefont-otf | fonts-freefont-ttf libavalon-framework-java  
  libcommons-logging-java-doc libexcalibur-logkit-java liblog4j1.2-java poppler-  
utils ghostscript  
  fonts-japanese-mincho | fonts-ipafont-mincho fonts-japanese-gothic | fonts-  
ipafont-gothic  
  fonts-arphic-ukai fonts-arphic-uming fonts-nanum ri ruby-dev bundler debhelper  
gv  
  | postscript-viewer perl-tk xpdf | pdf-viewer xzdec texlive-fonts-recommended-
```

```

doc
  texlive-latex-base-doc python3-pygments icc-profiles libfile-which-perl
  libspreadsheet-parseexcel-perl texlive-latex-extra-doc texlive-latex-
recommended-doc
  texlive-luatex texlive-pstricks dot2tex prerex texlive-pictures-doc vprerex
default-jre-headless
  tipa-doc
The following NEW packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-
texgyre
  fonts-urw-base35 libapache-pom-java libcommons-logging-java libcommons-parent-
java
  libfontbox-java libfontenc1 libgs9 libgs9-common libidn12 libijs-0.35
libjbig2dec0 libkpathsea6
  libpdfbox-java libptexenc1 libruby3.0 libsyntaxtex2 libteckit0 libtexlua53
libtexluajit2 libwoff1
  libzzip-0-13 lmodern poppler-data preview-latex-style rake ruby ruby-net-
telnet ruby-rubygems
  ruby-webrick ruby-xmlrpc ruby3.0 rubygems-integration t1utils teckit tex-
common tex-gyre texlive
  texlive-base texlive-binaries texlive-fonts-recommended texlive-latex-base
texlive-latex-extra
  texlive-latex-recommended texlive-pictures texlive-plain-generic texlive-xetex
tipa
  xfonts-encodings xfonts-utils
0 upgraded, 55 newly installed, 0 to remove and 35 not upgraded.
Need to get 182 MB of archives.
After this operation, 572 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-droid-fallback all
1:6.0.1r16-1.1build1 [1,805 kB]
Get:2 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-lato all 2.0-2.1
[2,696 kB]
Get:3 http://archive.ubuntu.com/ubuntu jammy/main amd64 poppler-data all
0.4.11-1 [2,171 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-common all 6.17
[33.7 kB]
Get:5 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-urw-base35 all
20200910-1 [6,367 kB]
Get:6 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9-common
all 9.55.0~dfsg1-0ubuntu5.6 [751 kB]
Get:7 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libidn12 amd64
1.38-4ubuntu1 [60.0 kB]
Get:8 http://archive.ubuntu.com/ubuntu jammy/main amd64 libijs-0.35 amd64
0.35-15build2 [16.5 kB]
Get:9 http://archive.ubuntu.com/ubuntu jammy/main amd64 libjbig2dec0 amd64
0.19-3build2 [64.7 kB]
Get:10 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9 amd64
9.55.0~dfsg1-0ubuntu5.6 [5,031 kB]

```

Get:11 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libkpathsea6 amd64 2021.20210626.59705-1ubuntu0.1 [60.3 kB]
Get:12 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libwoff1 amd64 1.0.2-1build4 [45.2 kB]
Get:13 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 dvisvgm amd64 2.13.1-1 [1,221 kB]
Get:14 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 fonts-lmodern all 2.004.5-6.1 [4,532 kB]
Get:15 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-noto-mono all 20201225-1build1 [397 kB]
Get:16 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 fonts-texgyre all 20180621-3.1 [10.2 MB]
Get:17 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libapache-pom-java all 18-1 [4,720 B]
Get:18 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcommons-parent-java all 43-1 [10.8 kB]
Get:19 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcommons-logging-java all 1.2-2 [60.3 kB]
Get:20 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libfontenc1 amd64 1:1.1.4-1build3 [14.7 kB]
Get:21 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libptexenc1 amd64 2021.20210626.59705-1ubuntu0.1 [39.1 kB]
Get:22 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 rubygems-integration all 1.18 [5,336 B]
Get:23 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby3.0 amd64 3.0.2-7ubuntu2.4 [50.1 kB]
Get:24 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby-rubygems all 3.3.5-2 [228 kB]
Get:25 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby amd64 1:3.0~exp1 [5,100 B]
Get:26 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 rake all 13.0.6-2 [61.7 kB]
Get:27 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby-net-telnet all 0.1.1-2 [12.6 kB]
Get:28 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 ruby-webrick all 1.7.0-3 [51.8 kB]
Get:29 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby-xmlrpc all 0.3.2-1ubuntu0.1 [24.9 kB]
Get:30 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libruby3.0 amd64 3.0.2-7ubuntu2.4 [5,113 kB]
Get:31 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libsynchronet2 amd64 2021.20210626.59705-1ubuntu0.1 [55.5 kB]
Get:32 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libteckit0 amd64 2.5.11+ds1-1 [421 kB]
Get:33 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libtexlua53 amd64 2021.20210626.59705-1ubuntu0.1 [120 kB]
Get:34 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libtexluajit2 amd64 2021.20210626.59705-1ubuntu0.1 [267 kB]

Get:35 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libzzip-0-13 amd64
 0.13.72+dfsg.1-1.1 [27.0 kB]
 Get:36 http://archive.ubuntu.com/ubuntu jammy/main amd64 xfonts-encodings all
 1:1.0.5-0ubuntu2 [578 kB]
 Get:37 http://archive.ubuntu.com/ubuntu jammy/main amd64 xfonts-utils amd64
 1:7.7+6build2 [94.6 kB]
 Get:38 http://archive.ubuntu.com/ubuntu jammy/universe amd64 lmodern all
 2.004.5-6.1 [9,471 kB]
 Get:39 http://archive.ubuntu.com/ubuntu jammy/universe amd64 preview-latex-style
 all 12.2-1ubuntu1 [185 kB]
 Get:40 http://archive.ubuntu.com/ubuntu jammy/main amd64 t1utils amd64
 1.41-4build2 [61.3 kB]
 Get:41 http://archive.ubuntu.com/ubuntu jammy/universe amd64 teckit amd64
 2.5.11+ds1-1 [699 kB]
 Get:42 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-gyre all
 20180621-3.1 [6,209 kB]
 Get:43 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 texlive-
 binaries amd64 2021.20210626.59705-1ubuntu0.1 [9,848 kB]
 Get:44 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-base all
 2021.20220204-1 [21.0 MB]
 Get:45 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-fonts-
 recommended all 2021.20220204-1 [4,972 kB]
 Get:46 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-base
 all 2021.20220204-1 [1,128 kB]
 Get:47 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-
 recommended all 2021.20220204-1 [14.4 MB]
 Get:48 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive all
 2021.20220204-1 [14.3 kB]
 Get:49 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libfontbox-java all
 1:1.8.16-2 [207 kB]
 Get:50 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libpdfbox-java all
 1:1.8.16-2 [5,199 kB]
 Get:51 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-pictures
 all 2021.20220204-1 [8,720 kB]
 Get:52 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-extra
 all 2021.20220204-1 [13.9 MB]
 Get:53 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-plain-
 generic all 2021.20220204-1 [27.5 MB]
 Get:54 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tipa all 2:1.3-21
 [2,967 kB]
 Get:55 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-xetex all
 2021.20220204-1 [12.4 MB]
 Fetched 182 MB in 16s (11.4 MB/s)
 Extracting templates from packages: 100%
 Preconfiguring packages ...
 Selecting previously unselected package fonts-droid-fallback.
 (Reading database ... 121749 files and directories currently installed.)
 Preparing to unpack .../00-fonts-droid-fallback_1%3a6.0.1r16-1.1build1_all.deb

```

...
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1build1) ...
Selecting previously unselected package fonts-lato.
Preparing to unpack .../01-fonts-lato_2.0-2.1_all.deb ...
Unpacking fonts-lato (2.0-2.1) ...
Selecting previously unselected package poppler-data.
Preparing to unpack .../02-poppler-data_0.4.11-1_all.deb ...
Unpacking poppler-data (0.4.11-1) ...
Selecting previously unselected package tex-common.
Preparing to unpack .../03-tex-common_6.17_all.deb ...
Unpacking tex-common (6.17) ...
Selecting previously unselected package fonts-urw-base35.
Preparing to unpack .../04-fonts-urw-base35_20200910-1_all.deb ...
Unpacking fonts-urw-base35 (20200910-1) ...
Selecting previously unselected package libgs9-common.
Preparing to unpack .../05-libgs9-common_9.55.0~dfsg1-0ubuntu5.6_all.deb ...
Unpacking libgs9-common (9.55.0~dfsg1-0ubuntu5.6) ...
Selecting previously unselected package libidn12:amd64.
Preparing to unpack .../06-libidn12_1.38-4ubuntu1_amd64.deb ...
Unpacking libidn12:amd64 (1.38-4ubuntu1) ...
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack .../07-libijs-0.35_0.35-15build2_amd64.deb ...
Unpacking libijs-0.35:amd64 (0.35-15build2) ...
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack .../08-libjbig2dec0_0.19-3build2_amd64.deb ...
Unpacking libjbig2dec0:amd64 (0.19-3build2) ...
Selecting previously unselected package libgs9:amd64.
Preparing to unpack .../09-libgs9_9.55.0~dfsg1-0ubuntu5.6_amd64.deb ...
Unpacking libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.6) ...
Selecting previously unselected package libkpathsea6:amd64.
Preparing to unpack .../10-libkpathsea6_2021.20210626.59705-1ubuntu0.1_amd64.deb
...
Unpacking libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Selecting previously unselected package libwoff1:amd64.
Preparing to unpack .../11-libwoff1_1.0.2-1build4_amd64.deb ...
Unpacking libwoff1:amd64 (1.0.2-1build4) ...
Selecting previously unselected package dvisvgm.
Preparing to unpack .../12-dvisvgm_2.13.1-1_amd64.deb ...
Unpacking dvisvgm (2.13.1-1) ...
Selecting previously unselected package fonts-lmodern.
Preparing to unpack .../13-fonts-lmodern_2.004.5-6.1_all.deb ...
Unpacking fonts-lmodern (2.004.5-6.1) ...
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack .../14-fonts-noto-mono_20201225-1build1_all.deb ...
Unpacking fonts-noto-mono (20201225-1build1) ...
Selecting previously unselected package fonts-texgyre.
Preparing to unpack .../15-fonts-texgyre_20180621-3.1_all.deb ...
Unpacking fonts-texgyre (20180621-3.1) ...

```

```

Selecting previously unselected package libapache-pom-java.
Preparing to unpack .../16-libapache-pom-java_18-1_all.deb ...
Unpacking libapache-pom-java (18-1) ...
Selecting previously unselected package libcommons-parent-java.
Preparing to unpack .../17-libcommons-parent-java_43-1_all.deb ...
Unpacking libcommons-parent-java (43-1) ...
Selecting previously unselected package libcommons-logging-java.
Preparing to unpack .../18-libcommons-logging-java_1.2-2_all.deb ...
Unpacking libcommons-logging-java (1.2-2) ...
Selecting previously unselected package libfontenc1:amd64.
Preparing to unpack .../19-libfontenc1_1%3a1.1.4-1build3_amd64.deb ...
Unpacking libfontenc1:amd64 (1:1.1.4-1build3) ...
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack .../20-libptexenc1_2021.20210626.59705-1ubuntu0.1_amd64.deb
...
Unpacking libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Selecting previously unselected package rubygems-integration.
Preparing to unpack .../21-rubygems-integration_1.18_all.deb ...
Unpacking rubygems-integration (1.18) ...
Selecting previously unselected package ruby3.0.
Preparing to unpack .../22-ruby3.0_3.0.2-7ubuntu2.4_amd64.deb ...
Unpacking ruby3.0 (3.0.2-7ubuntu2.4) ...
Selecting previously unselected package ruby-rubygems.
Preparing to unpack .../23-ruby-rubygems_3.3.5-2_all.deb ...
Unpacking ruby-rubygems (3.3.5-2) ...
Selecting previously unselected package ruby.
Preparing to unpack .../24-ruby_1%3a3.0~exp1_amd64.deb ...
Unpacking ruby (1:3.0~exp1) ...
Selecting previously unselected package rake.
Preparing to unpack .../25-rake_13.0.6-2_all.deb ...
Unpacking rake (13.0.6-2) ...
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack .../26-ruby-net-telnet_0.1.1-2_all.deb ...
Unpacking ruby-net-telnet (0.1.1-2) ...
Selecting previously unselected package ruby-webrick.
Preparing to unpack .../27-ruby-webrick_1.7.0-3_all.deb ...
Unpacking ruby-webrick (1.7.0-3) ...
Selecting previously unselected package ruby-xmlrpc.
Preparing to unpack .../28-ruby-xmlrpc_0.3.2-1ubuntu0.1_all.deb ...
Unpacking ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Selecting previously unselected package libruby3.0:amd64.
Preparing to unpack .../29-libruby3.0_3.0.2-7ubuntu2.4_amd64.deb ...
Unpacking libruby3.0:amd64 (3.0.2-7ubuntu2.4) ...
Selecting previously unselected package libsyntax2:amd64.
Preparing to unpack .../30-libsyntax2_2021.20210626.59705-1ubuntu0.1_amd64.deb
...
Unpacking libsyntax2:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Selecting previously unselected package libteckit0:amd64.

```



```

Preparing to unpack .../31-libteckit0_2.5.11+ds1-1_amd64.deb ...
Unpacking libteckit0:amd64 (2.5.11+ds1-1) ...
Selecting previously unselected package libtexlua53:amd64.
Preparing to unpack .../32-libtexlua53_2021.20210626.59705-1ubuntu0.1_amd64.deb
...
Unpacking libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Selecting previously unselected package libtexluajit2:amd64.
Preparing to unpack
.../33-libtexluajit2_2021.20210626.59705-1ubuntu0.1_amd64.deb ...
Unpacking libtexluajit2:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Selecting previously unselected package libzip-0-13:amd64.
Preparing to unpack .../34-libzip-0-13_0.13.72+dfsg.1-1.1_amd64.deb ...
Unpacking libzip-0-13:amd64 (0.13.72+dfsg.1-1.1) ...
Selecting previously unselected package xfonts-encodings.
Preparing to unpack .../35-xfonts-encodings_1%3a1.0.5-0ubuntu2_all.deb ...
Unpacking xfonts-encodings (1:1.0.5-0ubuntu2) ...
Selecting previously unselected package xfonts-utils.
Preparing to unpack .../36-xfonts-utils_1%3a7.7+6build2_amd64.deb ...
Unpacking xfonts-utils (1:7.7+6build2) ...
Selecting previously unselected package lmodern.
Preparing to unpack .../37-lmodern_2.004.5-6.1_all.deb ...
Unpacking lmodern (2.004.5-6.1) ...
Selecting previously unselected package preview-latex-style.
Preparing to unpack .../38-preview-latex-style_12.2-1ubuntu1_all.deb ...
Unpacking preview-latex-style (12.2-1ubuntu1) ...
Selecting previously unselected package t1utils.
Preparing to unpack .../39-t1utils_1.41-4build2_amd64.deb ...
Unpacking t1utils (1.41-4build2) ...
Selecting previously unselected package teckit.
Preparing to unpack .../40-teckit_2.5.11+ds1-1_amd64.deb ...
Unpacking teckit (2.5.11+ds1-1) ...
Selecting previously unselected package tex-gyre.
Preparing to unpack .../41-tex-gyre_20180621-3.1_all.deb ...
Unpacking tex-gyre (20180621-3.1) ...
Selecting previously unselected package texlive-binaries.
Preparing to unpack .../42-texlive-
binaries_2021.20210626.59705-1ubuntu0.1_amd64.deb ...
Unpacking texlive-binaries (2021.20210626.59705-1ubuntu0.1) ...
Selecting previously unselected package texlive-base.
Preparing to unpack .../43-texlive-base_2021.20220204-1_all.deb ...
Unpacking texlive-base (2021.20220204-1) ...
Selecting previously unselected package texlive-fonts-recommended.
Preparing to unpack .../44-texlive-fonts-recommended_2021.20220204-1_all.deb ...
Unpacking texlive-fonts-recommended (2021.20220204-1) ...
Selecting previously unselected package texlive-latex-base.
Preparing to unpack .../45-texlive-latex-base_2021.20220204-1_all.deb ...
Unpacking texlive-latex-base (2021.20220204-1) ...
Selecting previously unselected package texlive-latex-recommended.

```

```

Preparing to unpack .../46-texlive-latex-recommended_2021.20220204-1_all.deb ...
Unpacking texlive-latex-recommended (2021.20220204-1) ...
Selecting previously unselected package texlive.
Preparing to unpack .../47-texlive_2021.20220204-1_all.deb ...
Unpacking texlive (2021.20220204-1) ...
Selecting previously unselected package libfontbox-java.
Preparing to unpack .../48-libfontbox-java_1%3a1.8.16-2_all.deb ...
Unpacking libfontbox-java (1:1.8.16-2) ...
Selecting previously unselected package libpdfbox-java.
Preparing to unpack .../49-libpdfbox-java_1%3a1.8.16-2_all.deb ...
Unpacking libpdfbox-java (1:1.8.16-2) ...
Selecting previously unselected package texlive-pictures.
Preparing to unpack .../50-texlive-pictures_2021.20220204-1_all.deb ...
Unpacking texlive-pictures (2021.20220204-1) ...
Selecting previously unselected package texlive-latex-extra.
Preparing to unpack .../51-texlive-latex-extra_2021.20220204-1_all.deb ...
Unpacking texlive-latex-extra (2021.20220204-1) ...
Selecting previously unselected package texlive-plain-generic.
Preparing to unpack .../52-texlive-plain-generic_2021.20220204-1_all.deb ...
Unpacking texlive-plain-generic (2021.20220204-1) ...
Selecting previously unselected package tipa.
Preparing to unpack .../53-tipa_2%3a1.3-21_all.deb ...
Unpacking tipa (2:1.3-21) ...
Selecting previously unselected package texlive-xetex.
Preparing to unpack .../54-texlive-xetex_2021.20220204-1_all.deb ...
Unpacking texlive-xetex (2021.20220204-1) ...
Setting up fonts-lato (2.0-2.1) ...
Setting up fonts-noto-mono (20201225-1build1) ...
Setting up libwoff1:amd64 (1.0.2-1build4) ...
Setting up libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Setting up libijs-0.35:amd64 (0.35-15build2) ...
Setting up libtexluaajit2:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Setting up libfontbox-java (1:1.8.16-2) ...
Setting up rubygems-integration (1.18) ...
Setting up libzip-0-13:amd64 (0.13.72+dfsg.1-1.1) ...
Setting up fonts-urw-base35 (20200910-1) ...
Setting up poppler-data (0.4.11-1) ...
Setting up tex-common (6.17) ...
update-language: texlive-base not installed and configured, doing nothing!
Setting up libfontenc1:amd64 (1:1.1.4-1build3) ...
Setting up libjbig2dec0:amd64 (0.19-3build2) ...
Setting up libteckit0:amd64 (2.5.11+ds1-1) ...
Setting up libapache-pom-java (18-1) ...
Setting up ruby-net-telnet (0.1.1-2) ...
Setting up xfonts-encodings (1:1.0.5-0ubuntu2) ...
Setting up t1utils (1.41-4build2) ...
Setting up libidn12:amd64 (1.38-4ubuntu1) ...
Setting up fonts-texgyre (20180621-3.1) ...

```

```

Setting up libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Setting up ruby-webrick (1.7.0-3) ...
Setting up fonts-lmodern (2.004.5-6.1) ...
Setting up fonts-droid-fallback (1:6.0.1r16-1.1build1) ...
Setting up ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Setting up libsynchronet2:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Setting up libgs9-common (9.55.0~dfsg1-0ubuntu5.6) ...
Setting up teckit (2.5.11+ds1-1) ...
Setting up libpdfbox-java (1:1.8.16-2) ...
Setting up libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.6) ...
Setting up preview-latex-style (12.2-1ubuntu1) ...
Setting up libcommons-parent-java (43-1) ...
Setting up dvisvgm (2.13.1-1) ...
Setting up libcommons-logging-java (1.2-2) ...
Setting up xfonts-utils (1:7.7+6build2) ...
Setting up libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Setting up texlive-binaries (2021.20210626.59705-1ubuntu0.1) ...
update-alternatives: using /usr/bin/xdvi-xaw to provide /usr/bin/xdvi.bin
(xdvi.bin) in auto mode
update-alternatives: using /usr/bin/bibtex.original to provide /usr/bin/bibtex
(bibtex) in auto mode
Setting up lmodern (2.004.5-6.1) ...
Setting up texlive-base (2021.20220204-1) ...
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
mktexlsr: Updating /var/lib/texmf/ls-R-TEXLIVEDIST...
mktexlsr: Updating /var/lib/texmf/ls-R-TEXMFMAIN...
mktexlsr: Updating /var/lib/texmf/ls-R...
mktexlsr: Done.
tl-paper: setting paper size for dvips to a4:
/var/lib/texmf/dvips/config/config-paper.ps
tl-paper: setting paper size for dvipdfmx to a4:
/var/lib/texmf/dvipdfmx/dvipdfmx-paper.cfg
tl-paper: setting paper size for xdvi to a4: /var/lib/texmf/xdvi/XDvi-paper
tl-paper: setting paper size for pdftex to a4: /var/lib/texmf/tex/generic/tex-
ini-files/pdftexconfig.tex
Setting up tex-gyre (20180621-3.1) ...
Setting up texlive-plain-generic (2021.20220204-1) ...
Setting up texlive-latex-base (2021.20220204-1) ...
Setting up texlive-latex-recommended (2021.20220204-1) ...
Setting up texlive-pictures (2021.20220204-1) ...
Setting up texlive-fonts-recommended (2021.20220204-1) ...
Setting up tipa (2:1.3-21) ...
Setting up texlive (2021.20220204-1) ...
Setting up texlive-latex-extra (2021.20220204-1) ...
Setting up texlive-xetex (2021.20220204-1) ...

```

```

Setting up rake (13.0.6-2) ...
Setting up libruby3.0:amd64 (3.0.2-7ubuntu2.4) ...
Setting up ruby3.0 (3.0.2-7ubuntu2.4) ...
Setting up ruby (1:3.0~exp1) ...
Setting up ruby-rubygems (3.3.5-2) ...
Processing triggers for man-db (2.10.2-1) ...
Processing triggers for fontconfig (2.13.1-4.2ubuntu5) ...
Processing triggers for libc-bin (2.35-0ubuntu3.4) ...
/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic link

Processing triggers for tex-common (6.17) ...
Running updmap-sys. This may take some time... done.
Running mktexlsr /var/lib/texmf ... done.
Building format(s) --all.
    This may take some time... done.
Collecting py pandoc
  Downloading py pandoc-1.13-py3-none-any.whl (21 kB)
Installing collected packages: py pandoc
Successfully installed py pandoc-1.13

```

Conclusions

This project successfully leveraged BERT pre-trained model to perform sentiment analysis on a dataset of movie reviews, categorized into positive and negative sentiments. Through meticulous preprocessing and fine-tuning of the pre-trained model, we tailored it to the complexities of sentiment classification in movie reviews, achieving a model capable of distinguishing between positive and negative reviews with remarkable accuracy.

The evaluation metrics demonstrates the model's effectiveness in sentiment classification:

- For negative reviews (label 0), the model achieved a precision of 0.94, recall of 0.90, and F1-score of 0.92.

- For positive reviews (label 1), the model showed a precision of 0.90, recall of 0.93, and F1-score of 0.91.

- Overall accuracy stood at 0.92 across the dataset of 5000 reviews.

- The macro and weighted averages for precision, recall, and F1-score all mirrored the model's robust performance at 0.92, further validating the model's capability in handling both positive and

negative sentiments effectively.

These results are indicative of the model's robustness and its ability to generalize well to unseen movie reviews, affirming the power of pre-trained models in extracting meaningful insights from textual data. The high levels of precision and recall across both classes demonstrate the model's balanced capability in accurately classifying sentiments, making it a valuable tool for analyzing movie reviews.

```
[4]: !jupyter nbconvert --to PDF "Scrip3.ipynb"
```

```
[NbConvertApp] Converting notebook Scrip3.ipynb to PDF
/usr/local/lib/python3.10/dist-packages/nbconvert/filters/datatypefilter.py:41:
UserWarning: Your element with mimetype(s) dict_keys(['text/html']) is not able
to be represented.
  warn(
/usr/local/lib/python3.10/dist-packages/nbconvert/filters/datatypefilter.py:41:
UserWarning: Your element with mimetype(s) dict_keys(['text/html']) is not able
to be represented.
  warn(
/usr/local/lib/python3.10/dist-packages/nbconvert/filters/datatypefilter.py:41:
UserWarning: Your element with mimetype(s) dict_keys(['text/html']) is not able
to be represented.
  warn(
/usr/local/lib/python3.10/dist-packages/nbconvert/filters/datatypefilter.py:41:
UserWarning: Your element with mimetype(s) dict_keys(['text/html']) is not able
to be represented.
  warn(
[NbConvertApp] Support files will be in Scrip3_files/
[NbConvertApp] Making directory ./Scrip3_files
[NbConvertApp] Making directory ./Scrip3_files
[NbConvertApp] Making directory ./Scrip3_files
[NbConvertApp] Making directory ./Scrip3_files
[NbConvertApp] Making directory ./Scrip3_files
[NbConvertApp] Writing 163474 bytes to notebook.tex
[NbConvertApp] Building PDF
Traceback (most recent call last):
  File "/usr/local/bin/jupyter-nbconvert", line 8, in <module>
    sys.exit(main())
  File "/usr/local/lib/python3.10/dist-packages/jupyter_core/application.py",
line 283, in launch_instance
    super().launch_instance(argv=argv, **kwargs)
  File "/usr/local/lib/python3.10/dist-
packages/traitlets/config/application.py", line 992, in launch_instance
    app.start()
  File "/usr/local/lib/python3.10/dist-packages/nbconvert/nbconvertapp.py", line
423, in start
    self.convert_notebooks()
  File "/usr/local/lib/python3.10/dist-packages/nbconvert/nbconvertapp.py", line
```

```

597, in convert_notebooks
    self.convert_single_notebook(notebook_filename)
File "/usr/local/lib/python3.10/dist-packages/nbconvert/nbconvertapp.py", line
560, in convert_single_notebook
    output, resources = self.export_single_notebook(
File "/usr/local/lib/python3.10/dist-packages/nbconvert/nbconvertapp.py", line
488, in export_single_notebook
    output, resources = self.exporter.from_filename(
File "/usr/local/lib/python3.10/dist-
packages/nbconvert/exporters/exporter.py", line 189, in from_filename
    return self.from_file(f, resources=resources, **kw)
File "/usr/local/lib/python3.10/dist-
packages/nbconvert/exporters/exporter.py", line 206, in from_file
    return self.from_notebook_node(
File "/usr/local/lib/python3.10/dist-packages/nbconvert/exporters/pdf.py",
line 194, in from_notebook_node
    self.run_latex(tex_file)
File "/usr/local/lib/python3.10/dist-packages/nbconvert/exporters/pdf.py",
line 164, in run_latex
    return self.run_command(
File "/usr/local/lib/python3.10/dist-packages/nbconvert/exporters/pdf.py",
line 111, in run_command
    raise OSError(
OSError: xelatex not found on PATH, if you have not installed xelatex you may
need to do so. Find further instructions at
https://nbconvert.readthedocs.io/en/latest/install.html#installing-tex.

```

```

[ ]: !sudo apt-get install texlive-xetex texlive-fonts-recommended
      ↪texlive-plain-generic

```

```

Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
  fonts-texgyre fonts-urw-base35 libapache-pom-java libcommons-logging-java
  libcommons-parent-java libfontbox-java libfontenc1 libgs9 libgs9-common
  libidn12 libijs-0.35 libjbig2dec0 libkpathsea6 libpdfbox-java libptexenc1
  libruby3.0 libsynchronet2 libteckit0 libtexlua53 libtexluajit2 libwoff1
  libzip-0-13 lmodern poppler-data preview-latex-style rake ruby
  ruby-net-telnet ruby-rubygems ruby-webrick ruby-xmlrpc ruby3.0
  rubygems-integration tlutils teckit tex-common tex-gyre texlive-base
  texlive-binaries texlive-latex-base texlive-latex-extra
  texlive-latex-recommended texlive-pictures tipa xfonts-encodings
  xfonts-utils
Suggested packages:
  fonts-noto fonts-freefont-otf | fonts-freefont-ttf libavalon-framework-java
  libcommons-logging-java-doc libexcalibur-logkit-java liblog4j1.2-java

```

```

poppler-utils ghostscript fonts-japanese-mincho | fonts-ipafont-mincho
fonts-japanese-gothic | fonts-ipafont-gothic fonts-arphic-ukai
fonts-arphic-uming fonts-nanum ri ruby-dev bundler debhelper gv
| postscript-viewer perl-tk xpdf | pdf-viewer xzdec
texlive-fonts-recommended-doc texlive-latex-base-doc python3-pygments
icc-profiles libfile-which-perl libspreadsheet-parseexcel-perl
texlive-latex-extra-doc texlive-latex-recommended-doc texlive-luatex
texlive-pstricks dot2tex prerex texlive-pictures-doc vprerex
default-jre-headless tipa-doc

```

The following NEW packages will be installed:

```

dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
fonts-texgyre fonts-urw-base35 libapache-pom-java libcommons-logging-java
libcommons-parent-java libfontbox-java libfontenc1 libgs9 libgs9-common
libidn12 libijs-0.35 libjbig2dec0 libkpathsea6 libpdfbox-java libptexenc1
libruby3.0 libsynchronet2 libteckit0 libtexlua53 libtexlua53-luatex libwoff1
libzip-0-13 lmodern poppler-data preview-latex-style rake ruby
ruby-net-telnet ruby-rubygems ruby-webrick ruby-xmlrpc ruby3.0
rubygems-integration tlutils teckit tex-common tex-gyre texlive-base
texlive-binaries texlive-fonts-recommended texlive-latex-base
texlive-latex-extra texlive-latex-recommended texlive-pictures
texlive-plain-generic texlive-xetex tipa xfonts-encodings xfonts-utils

```

0 upgraded, 54 newly installed, 0 to remove and 35 not upgraded.

Need to get 182 MB of archives.

After this operation, 571 MB of additional disk space will be used.

Get:1 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-droid-fallback all 1:6.0.1r16-1.1build1 [1,805 kB]

Get:2 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-lato all 2.0-2.1 [2,696 kB]

Get:3 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 poppler-data all 0.4.11-1 [2,171 kB]

Get:4 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 tex-common all 6.17 [33.7 kB]

Get:5 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-urw-base35 all 20200910-1 [6,367 kB]

Get:6 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libgs9-common all 9.55.0~dfsg1-0ubuntu5.6 [751 kB]

Get:7 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libidn12 amd64 1.38-4ubuntu1 [60.0 kB]

Get:8 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libijs-0.35 amd64 0.35-15build2 [16.5 kB]

Get:9 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libjbig2dec0 amd64 0.19-3build2 [64.7 kB]

Get:10 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libgs9 amd64 9.55.0~dfsg1-0ubuntu5.6 [5,031 kB]

Get:11 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libkpathsea6 amd64 2021.20210626.59705-1ubuntu0.1 [60.3 kB]

Get:12 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libwoff1 amd64 1.0.2-1build4 [45.2 kB]

Get:13 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 dvisvgm amd64 2.13.1-1 [1,221 kB]
 Get:14 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 fonts-lmodern all 2.004.5-6.1 [4,532 kB]
 Get:15 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-noto-mono all 20201225-1build1 [397 kB]
 Get:16 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 fonts-texgyre all 20180621-3.1 [10.2 MB]
 Get:17 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libapache-pom-java all 18-1 [4,720 B]
 Get:18 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcommons-parent-java all 43-1 [10.8 kB]
 Get:19 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcommons-logging-java all 1.2-2 [60.3 kB]
 Get:20 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libfontenc1 amd64 1:1.1.4-1build3 [14.7 kB]
 Get:21 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libptexenc1 amd64 2021.20210626.59705-1ubuntu0.1 [39.1 kB]
 Get:22 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 rubygems-integration all 1.18 [5,336 B]
 Get:23 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby3.0 amd64 3.0.2-7ubuntu2.4 [50.1 kB]
 Get:24 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby-rubygems all 3.3.5-2 [228 kB]
 Get:25 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby amd64 1:3.0~exp1 [5,100 B]
 Get:26 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 rake all 13.0.6-2 [61.7 kB]
 Get:27 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby-net-telnet all 0.1.1-2 [12.6 kB]
 Get:28 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 ruby-webrick all 1.7.0-3 [51.8 kB]
 Get:29 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby-xmlrpc all 0.3.2-1ubuntu0.1 [24.9 kB]
 Get:30 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libruby3.0 amd64 3.0.2-7ubuntu2.4 [5,113 kB]
 Get:31 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libsinctex2 amd64 2021.20210626.59705-1ubuntu0.1 [55.5 kB]
 Get:32 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libteckit0 amd64 2.5.11+ds1-1 [421 kB]
 Get:33 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libtexlua53 amd64 2021.20210626.59705-1ubuntu0.1 [120 kB]
 Get:34 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libtexluajit2 amd64 2021.20210626.59705-1ubuntu0.1 [267 kB]
 Get:35 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libzip-0-13 amd64 0.13.72+dfsg.1-1.1 [27.0 kB]
 Get:36 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 xfonts-encodings all 1:1.0.5-0ubuntu2 [578 kB]


```

Get:37 http://archive.ubuntu.com/ubuntu jammy/main amd64 xfonts-utils amd64
1:7.7+6build2 [94.6 kB]
Get:38 http://archive.ubuntu.com/ubuntu jammy/universe amd64 lmodern all
2.004.5-6.1 [9,471 kB]
Get:39 http://archive.ubuntu.com/ubuntu jammy/universe amd64 preview-latex-style
all 12.2-1ubuntu1 [185 kB]
Get:40 http://archive.ubuntu.com/ubuntu jammy/main amd64 t1utils amd64
1.41-4build2 [61.3 kB]
Get:41 http://archive.ubuntu.com/ubuntu jammy/universe amd64 teckit amd64
2.5.11+ds1-1 [699 kB]
Get:42 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-gyre all
20180621-3.1 [6,209 kB]
Get:43 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 texlive-
binaries amd64 2021.20210626.59705-1ubuntu0.1 [9,848 kB]
Get:44 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-base all
2021.20220204-1 [21.0 MB]
Get:45 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-fonts-
recommended all 2021.20220204-1 [4,972 kB]
Get:46 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-base
all 2021.20220204-1 [1,128 kB]
Get:47 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libfontbox-java all
1:1.8.16-2 [207 kB]
Get:48 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libpdfbox-java all
1:1.8.16-2 [5,199 kB]
Get:49 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-
recommended all 2021.20220204-1 [14.4 MB]
Get:50 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-pictures
all 2021.20220204-1 [8,720 kB]
Get:51 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-extra
all 2021.20220204-1 [13.9 MB]
Get:52 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-plain-
generic all 2021.20220204-1 [27.5 MB]
Get:53 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tipa all 2:1.3-21
[2,967 kB]
Get:54 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-xetex all
2021.20220204-1 [12.4 MB]
Fetched 182 MB in 5s (35.7 MB/s)
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based
frontend cannot be used. at /usr/share/perl5/Debconf/FrontEnd/Dialog.pm line 78,
<> line 54.)
debconf: falling back to frontend: Readline
debconf: unable to initialize frontend: Readline
debconf: (This frontend requires a controlling tty.)
debconf: falling back to frontend: Teletype
dpkg-preconfigure: unable to re-open stdin:
Selecting previously unselected package fonts-droid-fallback.
(Reading database ... 121749 files and directories currently installed.)

```

```

Preparing to unpack .../00-fonts-droid-fallback_1%3a6.0.1r16-1.1build1_all.deb
...
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1build1) ...
Selecting previously unselected package fonts-lato.
Preparing to unpack .../01-fonts-lato_2.0-2.1_all.deb ...
Unpacking fonts-lato (2.0-2.1) ...
Selecting previously unselected package poppler-data.
Preparing to unpack .../02-poppler-data_0.4.11-1_all.deb ...
Unpacking poppler-data (0.4.11-1) ...
Selecting previously unselected package tex-common.
Preparing to unpack .../03-tex-common_6.17_all.deb ...
Unpacking tex-common (6.17) ...
Selecting previously unselected package fonts-urw-base35.
Preparing to unpack .../04-fonts-urw-base35_20200910-1_all.deb ...
Unpacking fonts-urw-base35 (20200910-1) ...
Selecting previously unselected package libgs9-common.
Preparing to unpack .../05-libgs9-common_9.55.0~dfsg1-0ubuntu5.6_all.deb ...
Unpacking libgs9-common (9.55.0~dfsg1-0ubuntu5.6) ...
Selecting previously unselected package libidn12:amd64.
Preparing to unpack .../06-libidn12_1.38-4ubuntu1_amd64.deb ...
Unpacking libidn12:amd64 (1.38-4ubuntu1) ...
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack .../07-libijs-0.35_0.35-15build2_amd64.deb ...
Unpacking libijs-0.35:amd64 (0.35-15build2) ...
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack .../08-libjbig2dec0_0.19-3build2_amd64.deb ...
Unpacking libjbig2dec0:amd64 (0.19-3build2) ...
Selecting previously unselected package libgs9:amd64.
Preparing to unpack .../09-libgs9_9.55.0~dfsg1-0ubuntu5.6_amd64.deb ...
Unpacking libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.6) ...
Selecting previously unselected package libkpathsea6:amd64.
Preparing to unpack .../10-libkpathsea6_2021.20210626.59705-1ubuntu0.1_amd64.deb
...
Unpacking libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Selecting previously unselected package libwoff1:amd64.
Preparing to unpack .../11-libwoff1_1.0.2-1build4_amd64.deb ...
Unpacking libwoff1:amd64 (1.0.2-1build4) ...
Selecting previously unselected package dvisvgm.
Preparing to unpack .../12-dvisvgm_2.13.1-1_amd64.deb ...
Unpacking dvisvgm (2.13.1-1) ...
Selecting previously unselected package fonts-lmodern.
Preparing to unpack .../13-fonts-lmodern_2.004.5-6.1_all.deb ...
Unpacking fonts-lmodern (2.004.5-6.1) ...
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack .../14-fonts-noto-mono_20201225-1build1_all.deb ...
Unpacking fonts-noto-mono (20201225-1build1) ...
Selecting previously unselected package fonts-texgyre.
Preparing to unpack .../15-fonts-texgyre_20180621-3.1_all.deb ...

```

```

Unpacking fonts-texgyre (20180621-3.1) ...
Selecting previously unselected package libapache-pom-java.
Preparing to unpack .../16-libapache-pom-java_18-1_all.deb ...
Unpacking libapache-pom-java (18-1) ...
Selecting previously unselected package libcommons-parent-java.
Preparing to unpack .../17-libcommons-parent-java_43-1_all.deb ...
Unpacking libcommons-parent-java (43-1) ...
Selecting previously unselected package libcommons-logging-java.
Preparing to unpack .../18-libcommons-logging-java_1.2-2_all.deb ...
Unpacking libcommons-logging-java (1.2-2) ...
Selecting previously unselected package libfontenc1:amd64.
Preparing to unpack .../19-libfontenc1_1%3a1.1.4-1build3_amd64.deb ...
Unpacking libfontenc1:amd64 (1:1.1.4-1build3) ...
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack .../20-libptexenc1_2021.20210626.59705-1ubuntu0.1_amd64.deb
...
Unpacking libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Selecting previously unselected package rubygems-integration.
Preparing to unpack .../21-rubygems-integration_1.18_all.deb ...
Unpacking rubygems-integration (1.18) ...
Selecting previously unselected package ruby3.0.
Preparing to unpack .../22-ruby3.0_3.0.2-7ubuntu2.4_amd64.deb ...
Unpacking ruby3.0 (3.0.2-7ubuntu2.4) ...
Selecting previously unselected package ruby-rubygems.
Preparing to unpack .../23-ruby-rubygems_3.3.5-2_all.deb ...
Unpacking ruby-rubygems (3.3.5-2) ...
Selecting previously unselected package ruby.
Preparing to unpack .../24-ruby_1%3a3.0~exp1_amd64.deb ...
Unpacking ruby (1:3.0~exp1) ...
Selecting previously unselected package rake.
Preparing to unpack .../25-rake_13.0.6-2_all.deb ...
Unpacking rake (13.0.6-2) ...
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack .../26-ruby-net-telnet_0.1.1-2_all.deb ...
Unpacking ruby-net-telnet (0.1.1-2) ...
Selecting previously unselected package ruby-webrick.
Preparing to unpack .../27-ruby-webrick_1.7.0-3_all.deb ...
Unpacking ruby-webrick (1.7.0-3) ...
Selecting previously unselected package ruby-xmlrpc.
Preparing to unpack .../28-ruby-xmlrpc_0.3.2-1ubuntu0.1_all.deb ...
Unpacking ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Selecting previously unselected package libruby3.0:amd64.
Preparing to unpack .../29-libruby3.0_3.0.2-7ubuntu2.4_amd64.deb ...
Unpacking libruby3.0:amd64 (3.0.2-7ubuntu2.4) ...
Selecting previously unselected package libsyntax2:amd64.
Preparing to unpack .../30-libsyntax2_2021.20210626.59705-1ubuntu0.1_amd64.deb
...
Unpacking libsyntax2:amd64 (2021.20210626.59705-1ubuntu0.1) ...

```

```

Selecting previously unselected package libteckit0:amd64.
Preparing to unpack .../31-libteckit0_2.5.11+ds1-1_amd64.deb ...
Unpacking libteckit0:amd64 (2.5.11+ds1-1) ...
Selecting previously unselected package libtexlua53:amd64.
Preparing to unpack .../32-libtexlua53_2021.20210626.59705-1ubuntu0.1_amd64.deb
...
Unpacking libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Selecting previously unselected package libtexluajit2:amd64.
Preparing to unpack
.../33-libtexluajit2_2021.20210626.59705-1ubuntu0.1_amd64.deb ...
Unpacking libtexluajit2:amd64 (2021.20210626.59705-1ubuntu0.1) ...
Selecting previously unselected package libzip-0-13:amd64.
Preparing to unpack .../34-libzip-0-13_0.13.72+dfsg.1-1.1_amd64.deb ...
Unpacking libzip-0-13:amd64 (0.13.72+dfsg.1-1.1) ...
Selecting previously unselected package xfonts-encodings.
Preparing to unpack .../35-xfonts-encodings_1%3a1.0.5-0ubuntu2_all.deb ...
Unpacking xfonts-encodings (1:1.0.5-0ubuntu2) ...
Selecting previously unselected package xfonts-utils.
Preparing to unpack .../36-xfonts-utils_1%3a7.7+6build2_amd64.deb ...
Unpacking xfonts-utils (1:7.7+6build2) ...
Selecting previously unselected package lmodern.
Preparing to unpack .../37-lmodern_2.004.5-6.1_all.deb ...
Unpacking lmodern (2.004.5-6.1) ...

```