

Predicting_Heart_Problem_with_BERT_in_Tensorflow Cleaned

5

July 30, 2020

##Mounting Google Drive

```
[49]: from google.colab import drive
      drive.mount("/GD")
```

Drive already mounted at /GD; to attempt to forcibly remount, call drive.mount("/GD", force_remount=True).

0.1 Importing Necessary Libraries

```
[50]: !pip install tensorflow==1.15.0
      import tensorflow as tf
      print(tf.__version__)
```

Requirement already satisfied: tensorflow==1.15.0 in /usr/local/lib/python3.6/dist-packages (1.15.0)
Requirement already satisfied: gast==0.2.2 in /usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (0.2.2)
Requirement already satisfied: opt-einsum>=2.3.2 in /usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (3.3.0)
Requirement already satisfied: protobuf>=3.6.1 in /usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (3.12.2)
Requirement already satisfied: six>=1.10.0 in /usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (1.15.0)
Requirement already satisfied: termcolor>=1.1.0 in /usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (1.1.0)
Requirement already satisfied: wheel>=0.26 in /usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (0.34.2)
Requirement already satisfied: absl-py>=0.7.0 in /usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (0.9.0)
Requirement already satisfied: tensorboard<1.16.0,>=1.15.0 in /usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (1.15.0)
Requirement already satisfied: keras-applications>=1.0.8 in /usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (1.0.8)
Requirement already satisfied: tensorflow-estimator==1.15.1 in /usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (1.15.1)
Requirement already satisfied: keras-preprocessing>=1.0.5 in

```

/usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (1.1.2)
Requirement already satisfied: wrapt>=1.11.1 in /usr/local/lib/python3.6/dist-
packages (from tensorflow==1.15.0) (1.12.1)
Requirement already satisfied: google-pasta>=0.1.6 in
/usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (0.2.0)
Requirement already satisfied: numpy<2.0,>=1.16.0 in
/usr/local/lib/python3.6/dist-packages (from tensorflow==1.15.0) (1.18.5)
Requirement already satisfied: grpcio>=1.8.6 in /usr/local/lib/python3.6/dist-
packages (from tensorflow==1.15.0) (1.30.0)
Requirement already satisfied: astor>=0.6.0 in /usr/local/lib/python3.6/dist-
packages (from tensorflow==1.15.0) (0.8.1)
Requirement already satisfied: setuptools in /usr/local/lib/python3.6/dist-
packages (from protobuf>=3.6.1->tensorflow==1.15.0) (49.1.0)
Requirement already satisfied: werkzeug>=0.11.15 in
/usr/local/lib/python3.6/dist-packages (from
tensorboard<1.16.0,>=1.15.0->tensorflow==1.15.0) (1.0.1)
Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.6/dist-
packages (from tensorboard<1.16.0,>=1.15.0->tensorflow==1.15.0) (3.2.2)
Requirement already satisfied: h5py in /usr/local/lib/python3.6/dist-packages
(from keras-applications>=1.0.8->tensorflow==1.15.0) (2.10.0)
Requirement already satisfied: importlib-metadata; python_version < "3.8" in
/usr/local/lib/python3.6/dist-packages (from
markdown>=2.6.8->tensorboard<1.16.0,>=1.15.0->tensorflow==1.15.0) (1.7.0)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.6/dist-
packages (from importlib-metadata; python_version <
"3.8"->markdown>=2.6.8->tensorboard<1.16.0,>=1.15.0->tensorflow==1.15.0) (3.1.0)
1.15.0

```

```

[51]: import pandas as pd
import tensorflow as tf
import tensorflow_hub as hub
from datetime import datetime
from sklearn.model_selection import train_test_split
import os

print("tensorflow version : ", tf.__version__)
print("tensorflow_hub version : ", hub.__version__)

```

```

tensorflow version : 1.15.0
tensorflow_hub version : 0.8.0

```

```

[56]: #Installing BERT module
!pip install bert-tensorflow

```

```

Requirement already satisfied: bert-tensorflow in /usr/local/lib/python3.6/dist-
packages (1.0.1)
Requirement already satisfied: six in /usr/local/lib/python3.6/dist-packages
(from bert-tensorflow) (1.15.0)

```

```
[57]: #Importing BERT modules
import bert
from bert import run_classifier
from bert import optimization
from bert import tokenization
```

0.2 ##Setting The Output Directory

While fine-tuning the model, we will save the training checkpoints and the model in an output directory so that we can use the trained model for our predictions later.

The following code block sets an output directory :

```
[58]: # Set the output directory for saving model file
OUTPUT_DIR = '/GD/My Drive/Colab Notebooks/LifeHackFinal/Clean'

#@markdown Whether or not to clear/delete the directory and create a new one
DO_DELETE = False #@param {type:"boolean"}

if DO_DELETE:
    try:
        tf.gfile.DeleteRecursively(OUTPUT_DIR)
    except:
        pass

tf.gfile.MakeDirs(OUTPUT_DIR)
print('***** Model output directory: {} *****'.format(OUTPUT_DIR))

***** Model output directory: /GD/My Drive/Colab Notebooks/LifeHackFinal/Clean
*****
```

0.3 ##Loading The Data

We will now load the data from a Google Drive directory and will also split the training set in to training and validation sets.

```
[60]: train = pd.read_csv("/GD/My Drive/Colab Notebooks/LifeHackFinal/Clean/
    ↪cleaned_sampled_data.csv", encoding = "ISO-8859-1")
train['question'] = train['question'].apply(str)

from sklearn.model_selection import train_test_split

train, val = train_test_split(train, test_size = 0.2, random_state = 100)

[61]: train.to_csv('trainFinal_cleaned.csv')
val.to_csv('testFinal_cleaned.csv')
```

```
[62]: #Training set sample
train.head(15)
```

```
[62]: Unnamed: 0 ... question
629      629 ... , sy mau . menderita hipertensi. sudah di ba...
669      669 ... , 1 bulan yang lalu saya periksa ke puskesmas ...
25       25 ... saya wanita berusia 24 tahu. sejak 5 tahun te...
88       88 ... dada trasa aga sesak tapi tidak batuk, 3hari ...
395     395 ... , sy pengidap jantung bocor, selalu kontrol se...
685     685 ... permisi, . saya berusia 21th, 3minggu yang lal...
1049    1049 ... , 6 bulan lalu saya terkena stroke (pecah pemb...
975     975 ... saya kaya terasa sesak terus kadang dada say...
356     356 ... bapak saya skrg lg d rawat di rs\r\nkepala pus...
146     146 ... ? denyut nadi bisa di pakai untuk menentuka...
1281    1281 ... , , ketika menjulurkan lidah, lidah saya domin...
45      45 ... , detakan jantung saya tiba-tiba kuat dan ter...
330     330 ... , tan mamah saya tangan dan kaki membengkak ...
160     160 ... , akhir" ini ketika saya menaiki tangga jantun...
306     306 ... ,, kira" umur 37th.. saring kaget kadang juga...
```

[15 rows x 5 columns]

```
[63]: print("Training Set Shape :", train.shape)
print("Validation Set Shape :", val.shape)
#print("Test Set Shape :", test.shape)
```

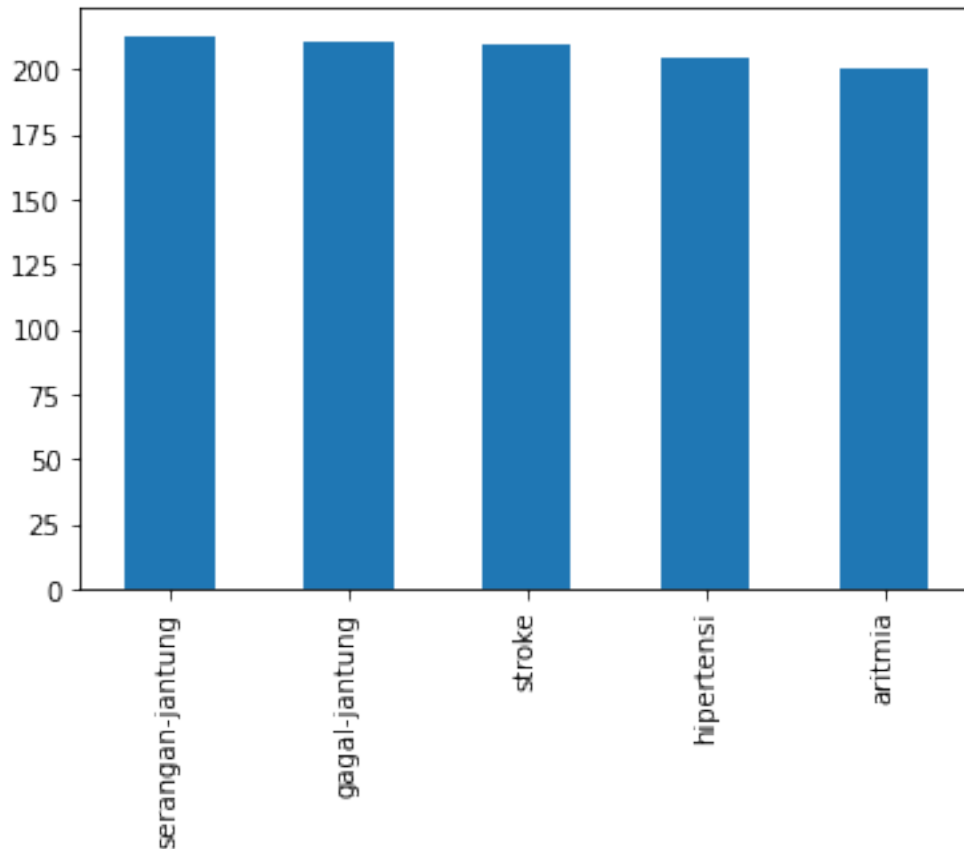
```
Training Set Shape : (1040, 5)
Validation Set Shape : (260, 5)
```

```
[64]: #unique classes
train['category'].unique()
```

```
[64]: array(['hipertensi', 'aritmia', 'gagal-jantung', 'stroke',
        'serangan-jantung'], dtype=object)
```

```
[65]: #Distribution of classes
train['category'].value_counts().plot(kind = 'bar')
```

```
[65]: <matplotlib.axes._subplots.AxesSubplot at 0x7f582f8fa0f0>
```



```
[66]: DATA_COLUMN = 'question'
      LABEL_COLUMN = 'category'
      # The list containing all the classes (train['SECTION'].unique())
      label_list = list(train['category'].unique())
```

0.4 Data Preprocessing

BERT model accept only a specific type of input and the datasets are usually structured to have the following four features:

- guid : A unique id that represents an observation.
- text_a : The text we need to classify into given categories
- text_b: It is used when we're training a model to understand the relationship between sentences and it does not apply for classification problems.
- label: It consists of the labels or classes or categories that a given text belongs to.

In our dataset we have text_a and label. The following code block will create objects for each of the above mentioned features for all the records in our dataset using the InputExample class provided in the BERT library.

```
[67]: train_InputExamples = train.apply(lambda x: bert.run_classifier.
      ↳ InputExample(guid=None,
      text_a =
      ↳ x[DATA_COLUMN],
      text_b =
      ↳ None,
      label =
      ↳ x[LABEL_COLUMN]), axis = 1)

val_InputExamples = val.apply(lambda x: bert.run_classifier.
      ↳ InputExample(guid=None,
      text_a =
      ↳ x[DATA_COLUMN],
      text_b =
      ↳ None,
      label =
      ↳ x[LABEL_COLUMN]), axis = 1)
```

```
[68]: train_InputExamples
```

```
[68]: 629 <bert.run_classifier.InputExample object at 0x...
      669 <bert.run_classifier.InputExample object at 0x...
      25 <bert.run_classifier.InputExample object at 0x...
      88 <bert.run_classifier.InputExample object at 0x...
      395 <bert.run_classifier.InputExample object at 0x...
      ...
      802 <bert.run_classifier.InputExample object at 0x...
      53 <bert.run_classifier.InputExample object at 0x...
      350 <bert.run_classifier.InputExample object at 0x...
      79 <bert.run_classifier.InputExample object at 0x...
      792 <bert.run_classifier.InputExample object at 0x...
      Length: 1040, dtype: object
```

```
[69]: print("Row 0 - guid of training set : ", train_InputExamples.iloc[0].guid)
      print("\n_____ \nRow 0 - text_a of training set : ", train_InputExamples.
      ↳ iloc[0].text_a)
      print("\n_____ \nRow 0 - text_b of training set : ", train_InputExamples.
      ↳ iloc[0].text_b)
      print("\n_____ \nRow 0 - label of training set : ", train_InputExamples.
      ↳ iloc[0].label)
```

Row 0 - guid of training set : None

 Row 0 - text_a of training set : , sy mau . menderita hipertensi. sudah di
 bawa ke dan diberi obat penurun hipertensi. tapi kok tekanan darahnya tidak
 kunjung normal? padahal sudah rutin minum obat & mengkonsumsi banyak buah2an

(timun, semangka, dll). malah badannya terasa lemas, pusing & penglihatan kabur.
terimakasih. w ð ð

Row 0 - text_b of training set : None

Row 0 - label of training set : hipertensi

We will now get down to business with the pretrained BERT. In this example we will use the bert_uncased_L-12_H-768_A-12/1 model. To check all available versions click [here](#).

We will be using the vocab.txt file in the model to map the words in the dataset to indexes. Also the loaded BERT model is trained on uncased/lowercase data and hence the data we feed to train the model should also be of lowercase.

The following code block loads the pre-trained BERT model and initializes a tokenizer object for tokenizing the texts.

```
[70]: # This is a path to an uncased (all lowercase) version of BERT
BERT_MODEL_HUB = "https://tfhub.dev/google/bert_multi_cased_L-12_H-768_A-12/1"

def create_tokenizer_from_hub_module():
    """Get the vocab file and casing info from the Hub module."""
    with tf.Graph().as_default():
        bert_module = hub.Module(BERT_MODEL_HUB)
        tokenization_info = bert_module(signature="tokenization_info", as_dict=True)
        with tf.Session() as sess:
            vocab_file, do_lower_case = sess.run([tokenization_info["vocab_file"],
                                                  tokenization_info["do_lower_case"]])

    return bert.tokenization.FullTokenizer(
        vocab_file=vocab_file, do_lower_case=do_lower_case)

tokenizer = create_tokenizer_from_hub_module()
```

INFO:tensorflow:Saver not created because there are no variables in the graph to restore

INFO:tensorflow:Saver not created because there are no variables in the graph to restore

```
[71]: #Here is what the tokenised sample of the first training set observation looks like
      ↪ like
print(tokenizer.tokenize(train_InputExamples.iloc[9].text_a))
```

```
['?', 'den', '##yu', '##t', 'nad', '##i', 'bisa', 'di', 'pak', '##ai', 'untuk',
'menentukan', 'be', '##ban', 'kerja', 'fi', '##sik', ',', 'tingkat', 'kes',
```

```
'##eh', '##atan', ',', 'tingkat', 'ke', '##bu', '##garan', 'dan', 'tingkat',  
'stress', '?']
```

We will now format out text in to input features which the BERT model expects. We will also set a sequence length which will be the length of the input features.

```
[72]: max_len = max([len(tokenizer.tokenize(train_InputExamples.iloc[IDX].text_a))  
    ↳ for IDX in range(1040)])  
print('Max length: ', max_len)
```

Max length: 2130

```
[73]: # We'll set sequences to be at most 128 tokens long.  
MAX_SEQ_LENGTH = 256  
  
# Convert our train and validation features to InputFeatures that BERT  
    ↳ understands.  
train_features = bert.run_classifier.  
    ↳ convert_examples_to_features(train_InputExamples, label_list,  
    ↳ MAX_SEQ_LENGTH, tokenizer)  
  
val_features = bert.run_classifier.  
    ↳ convert_examples_to_features(val_InputExamples, label_list, MAX_SEQ_LENGTH,  
    ↳ tokenizer)
```

INFO:tensorflow:Writing example 0 of 1040

INFO:tensorflow:Writing example 0 of 1040

INFO:tensorflow:*** Example ***

INFO:tensorflow:*** Example ***

INFO:tensorflow:guid: None

INFO:tensorflow:guid: None

INFO:tensorflow:tokens: [CLS] , sy mau . men ##der ##ita hip ##erten ##si .
sudah di ba ##wa ke dan diberi oba ##t pen ##uru ##n hip ##erten ##si . tapi ko
##k tekanan darah ##nya tidak kun ##jung normal ? pada ##hal sudah ru ##tin min
##um oba ##t & men ##g ##kon ##sum ##si banyak buah ##2 ##an (tim ##un , sem
##ang ##ka , dl ##l) . malah badan ##nya ter ##asa lema ##s , pus ##ing & pen
##gli ##hat ##an ka ##bur . ter ##ima ##kasi ##h . w ð ##ð [SEP]

INFO:tensorflow:tokens: [CLS] , sy mau . men ##der ##ita hip ##erten ##si .
sudah di ba ##wa ke dan diberi oba ##t pen ##uru ##n hip ##erten ##si . tapi ko
##k tekanan darah ##nya tidak kun ##jung normal ? pada ##hal sudah ru ##tin min
##um oba ##t & men ##g ##kon ##sum ##si banyak buah ##2 ##an (tim ##un , sem
##ang ##ka , dl ##l) . malah badan ##nya ter ##asa lema ##s , pus ##ing & pen
##gli ##hat ##an ka ##bur . ter ##ima ##kasi ##h . w ð ##ð [SEP]

INFO:tensorflow:input_ids: 101 117 12261 43024 119 10588 11304 11622 25377 26645
10449 119 25147 10120 15688 11037 11163 10215 50479 35355 10123 66558 25279
10115 25377 26645 10449 119 64747 11252 10174 93131 43947 10676 11868 13158
30425 16626 136 10585 18453 25147 13483 15364 13484 10465 35355 10123 111 10588
10240 17423 31417 10449 15175 21988 10729 10206 113 19604 11107 117 11531 11889
10371 117 63940 10161 114 119 73682 51463 10676 12718 23031 93661 10107 117
46960 10230 111 66558 20986 19180 10206 10730 34660 119 12718 12443 37997 10237
119 191 270 12332 102 0
0
0
0
0 0 0 0 0 0 0 0 0 0

INFO:tensorflow:input_ids: 101 117 12261 43024 119 10588 11304 11622 25377 26645
10449 119 25147 10120 15688 11037 11163 10215 50479 35355 10123 66558 25279
10115 25377 26645 10449 119 64747 11252 10174 93131 43947 10676 11868 13158
30425 16626 136 10585 18453 25147 13483 15364 13484 10465 35355 10123 111 10588
10240 17423 31417 10449 15175 21988 10729 10206 113 19604 11107 117 11531 11889
10371 117 63940 10161 114 119 73682 51463 10676 12718 23031 93661 10107 117
46960 10230 111 66558 20986 19180 10206 10730 34660 119 12718 12443 37997 10237
119 191 270 12332 102 0
0
0
0
0 0 0 0 0 0 0 0 0 0

INFO:tensorflow:input_mask: 1
1
1
0
0
0
0
0 0 0 0 0 0 0 0 0 0

INFO:tensorflow:input_mask: 1
1
1
0
0
0
0
0 0

INFO:tensorflow:segment_ids: 0
0
0
0
0
0
0 0

INFO:tensorflow:input_ids: 101 13935 117 119 11163 65899 10115 34675 22740 57236
15786 10671 42060 31877 119 15566 15780 11499 10526 16444 16184 10116 193 10240
25453 129 12385 32500 119 12591 21129 119 12430 29095 64981 15290 11015 76898
10174 10120 10141 14206 24091 119 37605 10120 56516 119 57236 24091 11910 53302
11249 119 64747 56331 10707 24091 20853 11163 57071 21550 10206 119 28344 29974
32310 45340 24091 10846 12257 11499 10526 29974 119 102 0 0 0 0 0 0 0 0 0 0 0 0
0
0
0
0
0
0 0 0 0 0 0 0 0

INFO:tensorflow:input_mask: 1
1
1 1 1 1 1 1 1 1 1 1 0
0
0
0
0
0 0

INFO:tensorflow:input_mask: 1
1
1 1 1 1 1 1 1 1 1 1 0
0
0
0
0
0 0

INFO:tensorflow:segment_ids: 0
0
0
0
0
0
0
0 0

INFO:tensorflow:segment_ids: 0
0
0
0
0
0
0
0 0

INFO:tensorflow:label: stroke (id = 3)

INFO:tensorflow:label: stroke (id = 3)

INFO:tensorflow:*** Example ***

INFO:tensorflow:*** Example ***

INFO:tensorflow:guid: None

INFO:tensorflow:guid: None

INFO:tensorflow:tokens: [CLS] . . apa ##kah orang y ##g mengalami sakit jan
##tung pada stadium 4 bisa be ##rku ##rang menjadi stadium 3 , stadium 3 menjadi
stadium 2 , stadium 2 menjadi stadium 1 , dan stadium 1 menjadi sem ##bu ##h
total ? jika bisa , bagaimana cara pen ##go ##batan ##nya ? moh ##on pen ##jela
##san ##nya . [SEP]

INFO:tensorflow:tokens: [CLS] . . apa ##kah orang y ##g mengalami sakit jan
##tung pada stadium 4 bisa be ##rku ##rang menjadi stadium 3 , stadium 3 menjadi
stadium 2 , stadium 2 menjadi stadium 1 , dan stadium 1 menjadi sem ##bu ##h
total ? jika bisa , bagaimana cara pen ##go ##batan ##nya ? moh ##on pen ##jela
##san ##nya . [SEP]

INFO:tensorflow:input_ids: 101 119 119 32500 28977 12430 193 10240 42060 57236
63923 23091 10585 27915 125 17103 10347 96315 24141 11999 27915 124 117 27915
124 11999 27915 123 117 27915 123 11999 27915 122 117 10215 27915 122 11999
11531 12177 10237 11339 136 37873 17103 117 100831 15903 66558 10797 47693 10676
136 49234 10263 66558 37142 14434 10676 119 102 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0
0
0
0
0 0

INFO:tensorflow:input_ids: 101 119 119 32500 28977 12430 193 10240 42060 57236
63923 23091 10585 27915 125 17103 10347 96315 24141 11999 27915 124 117 27915
124 11999 27915 123 117 27915 123 11999 27915 122 117 10215 27915 122 11999
11531 12177 10237 11339 136 37873 17103 117 100831 15903 66558 10797 47693 10676
136 49234 10263 66558 37142 14434 10676 119 102 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0
0
0
0
0 0

INFO:tensorflow:input_mask: 1
1 0 0 0 0
0
0
0
0
0 0

INFO:tensorflow:input_mask: 1
1 0 0 0 0
0
0
0 0

[illegible][illegible]

```
INFO:tensorflow:label: hipertensi (id = 0)
```

```
INFO:tensorflow:*** Example ***
```

INFO:tensorflow:guid: None

INFO:tensorflow:tokens: [CLS] , ab ##ang saya baru saja meninggal 3 hari yang lalu . pen ##ye ##bab ##nya kata yang mera ##wat adalah gagal jan ##tung . nah, . kira - kira penyakit jan ##tung yang ia alam ##i apa ##kah ada hubungan ##nya dengan penyakit seperti masuk angin , bu ##ang air dan lain - lain ? sia ##ng itu me ##mang dia men ##gel ##uhkan kepala pus ##ing , bu ##ang air , mun ##tah , dan dada ses ##ak . kami kira dia masuk angin , jadi kami kas ##ih tola ##k angin dan lain - lain . kemudian dia men ##gel ##uhkan , kala ##u tangan ##nya itu dan ja ##ri - ja ##rin ##ya sul ##it dig ##era ##kkan (dia be ##rp ##iki ##r ken ##astroke) . karena masih mua ##l - mua ##l , saya akhirnya pergi

untuk amb ##il salon ##pas , dan teman saya men ##jaga ab ##ang saya . lalu saya pulang , ab ##ang saya sudah tidak ada . ka sebelum meninggal itu , dia ke ##jang - ke ##jang luar biasa . sudah dibawa di rumah sakit , dilakukan c ##pr , namun tidak ter ##sel ##amat ##kan . beliau me ##mang memiliki ri ##way ##at darah rendah . dari semua ini , apa ##kah ada pen ##jela ##san yang saling berkaitan antara ge ##jala - ge ##jala yang disebut ##kan dia ##tas ? . . [SEP]

INFO:tensorflow:tokens: [CLS] , ab ##ang saya baru saja meninggal 3 hari yang lalu . pen ##ye ##bab ##nya kata yang mera ##wat adalah gagal jan ##tung . nah , . kira - kira penyakit jan ##tung yang ia alam ##i apa ##kah ada hubungan ##nya dengan penyakit seperti masuk angin , bu ##ang air dan lain - lain ? sia ##ng itu me ##mang dia men ##gel ##uhkan kepala pus ##ing , bu ##ang air , mun ##tah , dan dada ses ##ak . kami kira dia masuk angin , jadi kami kas ##ih tola ##k angin dan lain - lain . kemudian dia men ##gel ##uhkan , kala ##u tangan ##nya itu dan ja ##ri - ja ##rin ##ya sul ##it dig ##era ##kkan (dia be ##rp ##iki ##r ken ##a stroke) . karena masih mua ##l - mua ##l , saya akhirnya pergi untuk amb ##il salon ##pas , dan teman saya men ##jaga ab ##ang saya . lalu saya pulang , ab ##ang saya sudah tidak ada . ka sebelum meninggal itu , dia ke ##jang - ke ##jang luar biasa . sudah dibawa di rumah sakit , dilakukan c ##pr , namun tidak ter ##sel ##amat ##kan . beliau me ##mang memiliki ri ##way ##at darah rendah . dari semua ini , apa ##kah ada pen ##jela ##san yang saling berkaitan antara ge ##jala - ge ##jala yang disebut ##kan dia ##tas ? . . [SEP]

INFO:tensorflow:input_ids: 101 117 11357 11889 64981 18049 44725 31585 124 18370 10265 31288 119 66558 12871 51382 10676 21907 10265 71959 33670 10784 70591 63923 23091 119 64770 117 119 32105 118 32105 64951 63923 23091 10265 12729 40796 10116 32500 28977 15290 41585 10676 10659 64951 13908 34675 105676 117 11499 11889 12566 10215 13514 118 13514 136 13687 10376 11910 10911 45306 10671 10588 16039 68637 46687 46960 10230 117 11499 11889 12566 117 101833 53538 117 10215 42020 10974 10710 119 64985 32105 10671 34675 105676 117 17760 64985 14399 13187 90470 10174 105676 10215 13514 118 13514 119 16113 10671 10588 16039 68637 117 84844 10138 48371 10676 11910 10215 10201 10401 118 10201 13778 10679 12037 10486 80592 12015 33928 113 10671 10347 33394 20897 10129 67680 10113 57071 114 119 15786 20535 56944 10161 118 56944 10161 117 64981 30448 59159 10782 10559 11030 61658 20084 117 10215 71476 64981 10588 55539 11357 11889 64981 119 31288 64981 107874 117 11357 11889 64981 25147 11868 15290 119 10730 23667 31585 11910 117 10671 11163 37445 118 11163 37445 27120 34384 119 25147 95118 10120 22740 57236 117 28920 171 52302 117 22736 11868 12718 12912 49158 10706 119 19876 10911 45306 13363 29956 14132 10526 43947 47102 119 10397 23367 10592 117 32500 28977 15290 66558 37142 14434 10265 109002 85783 15345 46503 30216 118 46503 30216 10265 21250 10706 10671 11390 136 119 119 102 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

INFO:tensorflow:input_ids: 101 117 11357 11889 64981 18049 44725 31585 124 18370 10265 31288 119 66558 12871 51382 10676 21907 10265 71959 33670 10784 70591 63923 23091 119 64770 117 119 32105 118 32105 64951 63923 23091 10265 12729 40796 10116 32500 28977 15290 41585 10676 10659 64951 13908 34675 105676 117 11499 11889 12566 10215 13514 118 13514 136 13687 10376 11910 10911 45306 10671 10588 16039 68637 46687 46960 10230 117 11499 11889 12566 117 101833 53538 117


```
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
```

```
-----
Segment IDs : [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
```

##Creating A Multi-Class Classifier Model

```
[75]: def create_model(is_predicting, input_ids, input_mask, segment_ids, labels,
                        num_labels):

    bert_module = hub.Module(
        BERT_MODEL_HUB,
        trainable=True)
    bert_inputs = dict(
        input_ids=input_ids,
        input_mask=input_mask,
        segment_ids=segment_ids)
    bert_outputs = bert_module(
        inputs=bert_inputs,
        signature="tokens",
        as_dict=True)

    # Use "pooled_output" for classification tasks on an entire sentence.
    # Use "sequence_outputs" for token-level output.
    output_layer = bert_outputs["pooled_output"]

    hidden_size = output_layer.shape[-1].value

    # Create our own layer to tune for politeness data.
    output_weights = tf.get_variable(
        "output_weights", [num_labels, hidden_size],
        initializer=tf.truncated_normal_initializer(stddev=0.02))

    output_bias = tf.get_variable(
        "output_bias", [num_labels], initializer=tf.zeros_initializer())
```

```

with tf.variable_scope("loss"):

    # Dropout helps prevent overfitting
    output_layer = tf.nn.dropout(output_layer, keep_prob=0.9)

    logits = tf.matmul(output_layer, output_weights, transpose_b=True)
    logits = tf.nn.bias_add(logits, output_bias)
    log_probs = tf.nn.log_softmax(logits, axis=-1)

    # Convert labels into one-hot encoding
    one_hot_labels = tf.one_hot(labels, depth=num_labels, dtype=tf.float32)

    predicted_labels = tf.squeeze(tf.argmax(log_probs, axis=-1, output_type=tf.
→int32))
    # If we're predicting, we want predicted labels and the probabilities.
    if is_predicting:
        return (predicted_labels, log_probs)

    # If we're train/eval, compute loss between predicted and actual label
    per_example_loss = -tf.reduce_sum(one_hot_labels * log_probs, axis=-1)
    loss = tf.reduce_mean(per_example_loss)
    return (loss, predicted_labels, log_probs)

```

[76]: *#A function that adapts our model to work for training, evaluation, and*
→prediction.

```

# model_fn_builder actually creates our model function
# using the passed parameters for num_labels, learning_rate, etc.
def model_fn_builder(num_labels, learning_rate, num_train_steps,
                     num_warmup_steps):
    """Returns `model_fn` closure for TPUEstimator."""
    def model_fn(features, labels, mode, params): # pylint:
→disable=unused-argument
        """The `model_fn` for TPUEstimator."""

        input_ids = features["input_ids"]
        input_mask = features["input_mask"]
        segment_ids = features["segment_ids"]
        label_ids = features["label_ids"]

        is_predicting = (mode == tf.estimator.ModeKeys.PREDICT)

        # TRAIN and EVAL
        if not is_predicting:

            (loss, predicted_labels, log_probs) = create_model(

```

```

        is_predicting, input_ids, input_mask, segment_ids, label_ids,
↪num_labels)

train_op = bert.optimization.create_optimizer(
    loss, learning_rate, num_train_steps, num_warmup_steps, use_tpu=False)

# Calculate evaluation metrics.
def metric_fn(label_ids, predicted_labels):
    accuracy = tf.metrics.accuracy(label_ids, predicted_labels)
    true_pos = tf.metrics.true_positives(
        label_ids,
        predicted_labels)
    true_neg = tf.metrics.true_negatives(
        label_ids,
        predicted_labels)
    false_pos = tf.metrics.false_positives(
        label_ids,
        predicted_labels)
    false_neg = tf.metrics.false_negatives(
        label_ids,
        predicted_labels)

    return {
        "eval_accuracy": accuracy,
        "true_positives": true_pos,
        "true_negatives": true_neg,
        "false_positives": false_pos,
        "false_negatives": false_neg
    }

eval_metrics = metric_fn(label_ids, predicted_labels)

if mode == tf.estimator.ModeKeys.TRAIN:
    return tf.estimator.EstimatorSpec(mode=mode,
        loss=loss,
        train_op=train_op)
else:
    return tf.estimator.EstimatorSpec(mode=mode,
        loss=loss,
        eval_metric_ops=eval_metrics)
else:
    (predicted_labels, log_probs) = create_model(
        is_predicting, input_ids, input_mask, segment_ids, label_ids,
↪num_labels)

    predictions = {
        'probabilities': log_probs,

```

```

        'labels': predicted_labels
    }
    return tf.estimator.EstimatorSpec(mode, predictions=predictions)

# Return the actual model function in the closure
    return model_fn

```

```

[77]: # Compute train and warmup steps from batch size
# These hyperparameters are copied from this colab notebook (https://colab.
↳ sandbox.google.com/github/tensorflow/tpu/blob/master/tools/colab/
↳ bert_finetuning_with_cloud_tpus.ipynb)
BATCH_SIZE = 16
LEARNING_RATE = 2e-5
NUM_TRAIN_EPOCHS = 10
# Warmup is a period of time where the learning rate is small and gradually
↳ increases--usually helps training.
WARMUP_PROPORTION = 0.1
# Model configs
SAVE_CHECKPOINTS_STEPS = 300
SAVE_SUMMARY_STEPS = 100

# Compute train and warmup steps from batch size
num_train_steps = int(len(train_features) / BATCH_SIZE * NUM_TRAIN_EPOCHS)
num_warmup_steps = int(num_train_steps * WARMUP_PROPORTION)

# Specify output directory and number of checkpoint steps to save
run_config = tf.estimator.RunConfig(
    model_dir=OUTPUT_DIR,
    save_summary_steps=SAVE_SUMMARY_STEPS,
    save_checkpoints_steps=SAVE_CHECKPOINTS_STEPS)

# Specify output directory and number of checkpoint steps to save
run_config = tf.estimator.RunConfig(
    model_dir=OUTPUT_DIR,
    save_summary_steps=SAVE_SUMMARY_STEPS,
    save_checkpoints_steps=SAVE_CHECKPOINTS_STEPS)

```

```

[78]: #Initializing the model and the estimator
model_fn = model_fn_builder(
    num_labels=len(label_list),
    learning_rate=LEARNING_RATE,
    num_train_steps=num_train_steps,
    num_warmup_steps=num_warmup_steps)

estimator = tf.estimator.Estimator(
    model_fn=model_fn,
    config=run_config,

```

```
params={"batch_size": BATCH_SIZE})
```

```
INFO:tensorflow:Using config: {'_model_dir': '/GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean', '_tf_random_seed': None, '_save_summary_steps':
100, '_save_checkpoints_steps': 300, '_save_checkpoints_secs': None,
'_session_config': allow_soft_placement: true
graph_options {
  rewrite_options {
    meta_optimizer_iterations: ONE
  }
}
, '_keep_checkpoint_max': 5, '_keep_checkpoint_every_n_hours': 10000,
'_log_step_count_steps': 100, '_train_distribute': None, '_device_fn': None,
'_protocol': None, '_eval_distribute': None, '_experimental_distribute': None,
'_experimental_max_worker_delay_secs': None, '_session_creation_timeout_secs':
7200, '_service': None, '_cluster_spec':
<tensorflow.python.training.server_lib.ClusterSpec object at 0x7f5880215240>,
'_task_type': 'worker', '_task_id': 0, '_global_id_in_cluster': 0, '_master':
'', '_evaluation_master': '', '_is_chief': True, '_num_ps_replicas': 0,
'_num_worker_replicas': 1}
```

```
INFO:tensorflow:Using config: {'_model_dir': '/GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean', '_tf_random_seed': None, '_save_summary_steps':
100, '_save_checkpoints_steps': 300, '_save_checkpoints_secs': None,
'_session_config': allow_soft_placement: true
graph_options {
  rewrite_options {
    meta_optimizer_iterations: ONE
  }
}
, '_keep_checkpoint_max': 5, '_keep_checkpoint_every_n_hours': 10000,
'_log_step_count_steps': 100, '_train_distribute': None, '_device_fn': None,
'_protocol': None, '_eval_distribute': None, '_experimental_distribute': None,
'_experimental_max_worker_delay_secs': None, '_session_creation_timeout_secs':
7200, '_service': None, '_cluster_spec':
<tensorflow.python.training.server_lib.ClusterSpec object at 0x7f5880215240>,
'_task_type': 'worker', '_task_id': 0, '_global_id_in_cluster': 0, '_master':
'', '_evaluation_master': '', '_is_chief': True, '_num_ps_replicas': 0,
'_num_worker_replicas': 1}
```

we will now create an input builder function that takes our training feature set (`train_features`) and produces a generator. This is a pretty standard design pattern for working with Tensorflow Estimators.

```
[79]: # Create an input function for training. drop_remainder = True for using TPUs.
train_input_fn = bert.run_classifier.input_fn_builder(
    features=train_features,
    seq_length=MAX_SEQ_LENGTH,
```

```

        is_training=True,
        drop_remainder=False)

# Create an input function for validating. drop_remainder = True for using TPUs.
val_input_fn = run_classifier.input_fn_builder(
    features=val_features,
    seq_length=MAX_SEQ_LENGTH,
    is_training=False,
    drop_remainder=False)

```

Training & Evaluating

```

[80]: #Training the model
print(f'Beginning Training!')
current_time = datetime.now()
estimator.train(input_fn=train_input_fn, max_steps=num_train_steps)
print("Training took time ", datetime.now() - current_time)

```

Beginning Training!

INFO:tensorflow:Calling model_fn.

INFO:tensorflow:Calling model_fn.

INFO:tensorflow:Saver not created because there are no variables in the graph to restore

INFO:tensorflow:Saver not created because there are no variables in the graph to restore

/usr/local/lib/python3.6/dist-

packages/tensorflow_core/python/framework/indexed_slices.py:424: UserWarning: Converting sparse IndexedSlices to a dense Tensor of unknown shape. This may consume a large amount of memory.

"Converting sparse IndexedSlices to a dense Tensor of unknown shape. "

INFO:tensorflow:Done calling model_fn.

INFO:tensorflow:Done calling model_fn.

INFO:tensorflow:Create CheckpointSaverHook.

INFO:tensorflow:Create CheckpointSaverHook.

INFO:tensorflow:Graph was finalized.

INFO:tensorflow:Graph was finalized.

INFO:tensorflow:Running local_init_op.

INFO:tensorflow:Running local_init_op.

INFO:tensorflow:Done running local_init_op.

INFO:tensorflow:Done running local_init_op.

```
INFO:tensorflow:Saving checkpoints for 0 into /GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean/model.ckpt.

INFO:tensorflow:Saving checkpoints for 0 into /GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean/model.ckpt.

INFO:tensorflow:loss = 1.6149194, step = 0
INFO:tensorflow:loss = 1.6149194, step = 0
INFO:tensorflow:global_step/sec: 0.90728
INFO:tensorflow:global_step/sec: 0.90728
INFO:tensorflow:loss = 0.86942625, step = 100 (110.229 sec)
INFO:tensorflow:loss = 0.86942625, step = 100 (110.229 sec)
INFO:tensorflow:global_step/sec: 1.05845
INFO:tensorflow:global_step/sec: 1.05845
INFO:tensorflow:loss = 0.9200649, step = 200 (94.469 sec)
INFO:tensorflow:loss = 0.9200649, step = 200 (94.469 sec)
INFO:tensorflow:Saving checkpoints for 300 into /GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean/model.ckpt.

INFO:tensorflow:Saving checkpoints for 300 into /GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean/model.ckpt.

INFO:tensorflow:global_step/sec: 0.813108
INFO:tensorflow:global_step/sec: 0.813108
INFO:tensorflow:loss = 0.6114168, step = 300 (122.983 sec)
INFO:tensorflow:loss = 0.6114168, step = 300 (122.983 sec)
INFO:tensorflow:global_step/sec: 1.05795
INFO:tensorflow:global_step/sec: 1.05795
INFO:tensorflow:loss = 0.075308524, step = 400 (94.523 sec)
INFO:tensorflow:loss = 0.075308524, step = 400 (94.523 sec)
INFO:tensorflow:global_step/sec: 1.05994
INFO:tensorflow:global_step/sec: 1.05994
INFO:tensorflow:loss = 0.031500816, step = 500 (94.346 sec)
INFO:tensorflow:loss = 0.031500816, step = 500 (94.346 sec)
INFO:tensorflow:Saving checkpoints for 600 into /GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean/model.ckpt.

INFO:tensorflow:Saving checkpoints for 600 into /GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean/model.ckpt.
```



```

INFO:tensorflow:global_step/sec: 0.821167
INFO:tensorflow:global_step/sec: 0.821167
INFO:tensorflow:loss = 0.021934424, step = 600 (121.776 sec)
INFO:tensorflow:loss = 0.021934424, step = 600 (121.776 sec)
INFO:tensorflow:Saving checkpoints for 650 into /GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean/model.ckpt.
INFO:tensorflow:Saving checkpoints for 650 into /GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean/model.ckpt.
INFO:tensorflow:Loss for final step: 0.25760937.
INFO:tensorflow:Loss for final step: 0.25760937.
Training took time 0:13:20.106616

```

```

[81]: #Evaluating the model with Validation set
eval_results = estimator.evaluate(input_fn=val_input_fn, steps=None)

```

```

INFO:tensorflow:Calling model_fn.
INFO:tensorflow:Calling model_fn.
INFO:tensorflow:Saver not created because there are no variables in the graph to
restore
INFO:tensorflow:Saver not created because there are no variables in the graph to
restore
/usr/local/lib/python3.6/dist-
packages/tensorflow_core/python/framework/indexed_slices.py:424: UserWarning:
Converting sparse IndexedSlices to a dense Tensor of unknown shape. This may
consume a large amount of memory.
    "Converting sparse IndexedSlices to a dense Tensor of unknown shape. "
INFO:tensorflow:Done calling model_fn.
INFO:tensorflow:Done calling model_fn.
INFO:tensorflow:Starting evaluation at 2020-07-29T18:44:15Z
INFO:tensorflow:Starting evaluation at 2020-07-29T18:44:15Z
INFO:tensorflow:Graph was finalized.
INFO:tensorflow:Graph was finalized.
INFO:tensorflow:Restoring parameters from /GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean/model.ckpt-650
INFO:tensorflow:Restoring parameters from /GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean/model.ckpt-650
INFO:tensorflow:Running local_init_op.

```

```

INFO:tensorflow:Running local_init_op.
INFO:tensorflow:Done running local_init_op.
INFO:tensorflow:Done running local_init_op.
INFO:tensorflow:Finished evaluation at 2020-07-29-18:45:20
INFO:tensorflow:Finished evaluation at 2020-07-29-18:45:20
INFO:tensorflow:Saving dict for global step 650: eval_accuracy = 0.75,
false_negatives = 10.0, false_positives = 9.0, global_step = 650, loss =
1.0498428, true_negatives = 46.0, true_positives = 195.0
INFO:tensorflow:Saving dict for global step 650: eval_accuracy = 0.75,
false_negatives = 10.0, false_positives = 9.0, global_step = 650, loss =
1.0498428, true_negatives = 46.0, true_positives = 195.0
INFO:tensorflow:Saving 'checkpoint_path' summary for global step 650: /GD/My
Drive/Colab Notebooks/LifeHackFinal/Clean/model.ckpt-650
INFO:tensorflow:Saving 'checkpoint_path' summary for global step 650: /GD/My
Drive/Colab Notebooks/LifeHackFinal/Clean/model.ckpt-650

```

```
[82]: eval_results
```

```
[82]: {'eval_accuracy': 0.75,
      'false_negatives': 10.0,
      'false_positives': 9.0,
      'global_step': 650,
      'loss': 1.0498428,
      'true_negatives': 46.0,
      'true_positives': 195.0}
```

```
[83]: predictions = estimator.predict(val_input_fn)
```

```
[84]: preds_result = []
      for prediction in predictions:
          preds_result.append((prediction['probabilities'], prediction['labels']))
```

```

INFO:tensorflow:Calling model_fn.
INFO:tensorflow:Calling model_fn.
INFO:tensorflow:Saver not created because there are no variables in the graph to
restore
INFO:tensorflow:Saver not created because there are no variables in the graph to
restore
INFO:tensorflow:Done calling model_fn.
INFO:tensorflow:Done calling model_fn.
INFO:tensorflow:Graph was finalized.

```

```
INFO:tensorflow:Graph was finalized.

INFO:tensorflow:Restoring parameters from /GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean/model.ckpt-650

INFO:tensorflow:Restoring parameters from /GD/My Drive/Colab
Notebooks/LifeHackFinal/Clean/model.ckpt-650

INFO:tensorflow:Running local_init_op.

INFO:tensorflow:Running local_init_op.

INFO:tensorflow:Done running local_init_op.

INFO:tensorflow:Done running local_init_op.
```

```
[85]: y_pred = list(map(lambda x: x[1], preds_result))
```

```
[86]: mapping = dict()

for i in range(len(label_list)):
    mapping[label_list[i]] = i

y_actual = list(map(lambda x: mapping[x], val['category'].tolist()))
```

```
[87]: from sklearn.metrics import confusion_matrix

confusion_matrix(y_actual, y_pred)
```

```
[87]: array([[46,  1,  3,  4,  1],
            [ 2, 38,  8,  2,  9],
            [ 1,  2, 37,  2,  7],
            [ 5,  0,  3, 42,  0],
            [ 2,  5,  9,  0, 31]])
```

```
[88]: val_pred = val.copy()
val_pred['pred'] = list(map(lambda x: label_list[x], y_pred))
val_pred.to_csv('prediction_final_raw.csv')
```

```
[89]: from sklearn.metrics import classification_report
print(classification_report(val_pred['category'], val_pred['pred']))
```

	precision	recall	f1-score	support
aritmia	0.83	0.64	0.72	59
gagal-jantung	0.62	0.76	0.68	49
hipertensi	0.82	0.84	0.83	55
serangan-jantung	0.65	0.66	0.65	47
stroke	0.84	0.84	0.84	50
accuracy			0.75	260

macro avg	0.75	0.75	0.74	260
weighted avg	0.76	0.75	0.75	260

```
[90]: val_pred.head()
```

```
[90]:      Unnamed: 0  ...      pred
1001      1001  ... serangan-jantung
1266      1266  ...   gagal-jantung
503       503  ...   gagal-jantung
756       756  ...      hipertensi
459       459  ...   gagal-jantung
```

```
[5 rows x 6 columns]
```

#Reference: Most of the code has been taken from the following resource:

- https://colab.research.google.com/github/google-research/bert/blob/master/predicting_movie_reviews_w