entiment-analysis-pretrained-model

July 23, 2025

0.0.1 Sentiment Analysis using pretrained model

[ ]:
```
# Run these commands in the terminal to install the required libraries:
# pip install pandas
# pip install transformers
# pip install openpyxl
```

[1]:
```
import pandas as pd
```

[2]:
```
# Open xlsx file
df = pd.read_excel('data/student_survey-jessica.xlsx',␣
 ↪sheet_name='before_after')
```

[4]:
```
df.shape[0]
```

[4]: 2005

[7]:
```
int((1 / 2005) * 100)
```

[7]: 0

[3]:
```
# # drop nas for feedback_challenge5
# df = df.dropna(subset=['feedback_challenge5'])
# # count nas
# print(df['feedback_book'].isna().sum())
# print(df['feedback_challenge5'].isna().sum())
```

[4]:
```
# Sample code
# from transformers import pipeline
# sentiment_pipeline = pipeline("sentiment-analysis")
# data = [" it is a simple challenge but in addition to this it leaves you a␣
 ↪good teaching that will serve to base you on a future It is a simple␣
 ↪challenge but one that leaves in teaching that can last forever", "No"]
# sentiment_pipeline(data)
# data = ["It was relatively simple, but since I don't have a profession␣
 ↪selected, the questions were complicated for me.", "It was so-so"]
# sentiment_pipeline(data)
```

[5]:
```
from transformers import pipeline
# gen_sentiment = pipeline(model='Seethal/sentiment_analysis_generic_dataset')
tweet_sentiment = pipeline(model='cardiffnlp/twitter-roberta-base-sentiment')
```

```
C:\Users\Sakuta\AppData\Roaming\Python\Python311\site-packages\tqdm\auto.py:21:
TqdmWarning: IProgress not found. Please update jupyter and ipywidgets. See
https://ipywidgets.readthedocs.io/en/stable/user_install.html
  from .autonotebook import tqdm as notebook_tqdm
All model checkpoint layers were used when initializing
TFRobertaForSequenceClassification.

All the layers of TFRobertaForSequenceClassification were initialized from the
model checkpoint at cardiffnlp/twitter-roberta-base-sentiment.
If your task is similar to the task the model of the checkpoint was trained on,
you can already use TFRobertaForSequenceClassification for predictions without
further training.
```

[ ]:
```
# import tensorflow as tf

# from transformers import AutoTokenizer, TFRobertaForSequenceClassification
# tokenizer = AutoTokenizer.from_pretrained("cardiffnlp/
 ↪twitter-roberta-base-sentiment")
# model = TFRobertaForSequenceClassification.from_pretrained("cardiffnlp/
 ↪twitter-roberta-base-sentiment")
```

[ ]:
```
# inputs = tokenizer("An so-so story.", return_tensors="tf")
# logits = model(**inputs).logits
# # logits
# predicted_class_id = int(tf.math.argmax(logits, axis=-1)[0])
# model.config.id2label[predicted_class_id]
```

[ ]:
```
# data = [
#     "I thought it would take less time to do it, but despite that I like to␣
 ↪make the story since in a certain way I also vent",
#     " No comment ...",
#     "Everything okay.",
#     "Not really.",
#     "No, all right.",
#     "No, it's just necessary for us to keep our future in mind",
#     " it's very good No",
#     "No, I loved it",
#     "I thought it was a very different challenge from the others but it was␣
 ↪with a little reflection.",
#     "None",
#     "0",
#     "Somewhat long",
```

```
#     "Very long, at one point it became tedious to me, and I stopped reasoning␣
 ↪everything.",
#     "Nothing in particular",
#     "I think I identify myself with the brave apple",
#     "Well, they ask me questions that I think I have answers, but the truth␣
 ↪is that I do not and it cost me a lot. Last semester despite having drafted␣
 ↪and submitted my purpose, I didn't feel it as a purpose. Rather, it felt␣
 ↪like something forced that I did just to leave the class and not take more␣
 ↪time away from the teacher. So here I had to do it again, but again, without␣
 ↪feeling it as a purpose really.",
#     "", # empty string is classified as neutral with 0.45 score for␣
 ↪cardiffnlp/twitter-roberta-base-sentiment model
# ]
# [{i:data[i], d['label'].replace('LABEL_0', 'negative').replace('LABEL_1',␣
 ↪'neutral').replace('LABEL_2', 'positive'):d['score']} for i, d in␣
 ↪enumerate(tweet_sentiment(data))]
```

[ ]:
```
# list(df['feedback_book'][:100])
```

[ ]:
```
# df['feedback_challenge5'][25:]
```

[128]:
```
# str(df['feedback_challenge5'][28]) == "nan"
```

[128]: True

[10]:
```
# feedback_book_sentiment_list = [{d['label'].replace('LABEL_0', 'negative').
 ↪replace('LABEL_1', 'neutral').replace('LABEL_2', 'positive'):d['score']} for␣
 ↪d in (tweet_sentiment(list(df['feedback_book'].astype(str))))]
```

[ ]:
```
# feedback_book_sentiment_list
```

[ ]:
```
# for i, feedback_book in enumerate(df['feedback_book'].astype(str)):
#     print(i+1, feedback_book)
#     ts = tweet_sentiment([feedback_book])
#     print(ts[0]['label'].replace('LABEL_0', 'negative').replace('LABEL_1',␣
 ↪'neutral').replace('LABEL_2', 'positive'), ts[0]['score'])
```

[ ]:
```
# feedback_book_sentiment_list
```

[79]:
```
# feedback_book_sentiment_labels = [list(d.items())[0][0] for d in␣
 ↪feedback_book_sentiment_list]
# feedback_book_sentiment_scores = [list(d.items())[0][1] for d in␣
 ↪feedback_book_sentiment_list]
```

[ ]:
```
# feedback_book_sentiment_labels
```

[ ]:
```
# feedback_book_sentiment_scores
```

[83]:
```
# df2 = df.copy()
```

[ ]:
```
# df2.head(3)
```

[85]:
```
# df2['feedback_book_sentiment'] = feedback_book_sentiment_labels
# df2['feedback_book_sentiment_confidence'] = feedback_book_sentiment_scores
```

[ ]:
```
# df2.head()
```

[129]:
```
df2 = df.copy()
df2.insert(12, 'feedback_book_sentiment', value=None)
df2.insert(13, 'feedback_book_sentiment_confidence', value=None)
df2.insert(15, 'feedback_challenge5_sentiment', value=None)
df2.insert(16, 'feedback_challenge5_sentiment_confidence', value=None)
# loop through each row in the df
for i, row in df.iterrows():
    # if i > 5:
    #     break
    print(f'{i + 1}')

    feedback_book_output = tweet_sentiment(str(df['feedback_book'][i]))
    feedback_book_sentiment_label = feedback_book_output[0]['label'].
 ↪replace('LABEL_0', 'Negative').replace('LABEL_1', 'Neutral').
 ↪replace('LABEL_2', 'Positive')
    feedback_book_sentiment_score = feedback_book_output[0]['score']

    feedback_challenge5_output =␣
 ↪tweet_sentiment(str(df['feedback_challenge5'][i]))
    feedback_challenge5_sentiment_label =␣
 ↪feedback_challenge5_output[0]['label'].replace('LABEL_0', 'Negative').
 ↪replace('LABEL_1', 'Neutral').replace('LABEL_2', 'Positive')
    feedback_challenge5_sentiment_score = feedback_challenge5_output[0]['score']

    # add the sentiment label and score to the df2
    df2.loc[i, 'feedback_book_sentiment'] = feedback_book_sentiment_label
    df2.loc[i, 'feedback_book_sentiment_confidence'] =␣
 ↪feedback_book_sentiment_score
    df2.loc[i, 'feedback_challenge5_sentiment'] =␣
 ↪feedback_challenge5_sentiment_label
    df2.loc[i, 'feedback_challenge5_sentiment_confidence'] =␣
 ↪feedback_challenge5_sentiment_score

    if str(df['feedback_book'][i]).strip() in ["", "-", "0", "nan"]:
        df2.loc[i, 'feedback_book_sentiment'] = '-'
```

```
        df2.loc[i, 'feedback_book_sentiment_confidence'] = '-'
    if str(df['feedback_challenge5'][i]).strip() in ["", "-", "0", "nan"]:
        df2.loc[i, 'feedback_challenge5_sentiment'] = '-'
        df2.loc[i, 'feedback_challenge5_sentiment_confidence'] = '-'
```

1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962

```
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
```

[130]:
```
df2
```

[130]:

	student_id	survey_month	survey_year	before_lifepurpose
0	AL02674306	September	2021	Neither agree or disagree
1	AL02684742	September	2021	Neither agree or disagree
2	AL02685313	June	2022	Totally agree
3	AL02685686	September	2021	Partially agree
4	AL02685806	September	2021	Neither agree or disagree
…	…	…	…	…
2000	AL03061637	September	2022	Totally disagree
2001	AL03061692	September	2022	Partially disagree
2002	AL03061944	September	2022	Totally agree
2003	AL03068090	September	2022	Totally agree
2004	AL04510630	September	2021	Partially disagree

	before_personalresources	before_motivation
0	Partially agree	Totally agree
1	Partially disagree	Partially disagree
2	Totally agree	Totally agree
3	Partially agree	Partially agree
4	Partially disagree	Partially agree
…	…	…
2000	Partially agree	Neither agree or disagree
2001	Neither agree or disagree	Neither agree or disagree
2002	Totally agree	Totally agree
2003	Totally agree	Totally agree
2004	Partially agree	Neither agree or disagree

	after_lifepurpose	after_personalresources
0	Neither agree or disagree	Partially agree
1	Partially disagree	Partially disagree
2	Totally agree	Totally agree
3	Totally agree	Totally agree
4	Partially agree	Neither agree or disagree
…	…	…
2000	Partially disagree	Partially agree
2001	Partially agree	Partially agree
2002	Totally agree	Totally agree
2003	Totally agree	Totally agree
2004	Partially disagree	Partially agree

	after_motivation	recommendation_book
0	Totally agree	Totally agree
1	Partially disagree	Neither agree or disagree
2	Totally agree	Totally agree
3	Totally agree	Totally agree
4	Partially agree	Neither agree or disagree
…	…	…
2000	Totally agree	Neither agree or disagree

```
2001              Totally agree              Totally agree
2002              Totally agree           Partially agree
2003              Totally agree              Totally agree
2004  Neither agree or disagree           Partially agree

      recommendation_challenge5
0                 Totally agree  \
1     Neither agree or disagree
2                 Totally agree
3                 Totally agree
4     Neither agree or disagree
…                           …
2000              Totally agree
2001           Partially agree
2002  Neither agree or disagree
2003              Totally agree
2004           Partially agree

                                          feedback_book
0      It is a story related to situations of many p…  \
1                                          Interesting
2                                   An inspiring story.
3     It seems like a very simple story, even for ch…
4                                          It is good
…                                                 …
2000  I thought it was a different but assertive exp…
2001  I think it's quite an interesting story and it…
2002  It is a great story because by using metaphors…
2003  I really liked the story because of the way yo…
2004        Quite good, because it gives instrospection

     feedback_book_sentiment feedback_book_sentiment_confidence
0                   Positive                           0.838734  \
1                   Positive                            0.73105
2                   Positive                           0.921101
3                   Positive                            0.61259
4                   Positive                           0.967371
…                      …                                …
2000                Neutral                           0.759637
2001                Positive                          0.944936
2002                Positive                          0.977118
2003                Positive                          0.978657
2004                Positive                          0.907301

                                    feedback_challenge5
0      it is a simple challenge but in addition to t…  \
1                                                  No
```

```
2            It helped me think in a more positive way.
3                                                  none
4                   I have no comment on the challenge.
…                                                   …
2000                                                  No
2001                                                  No
2002  It is a bit of a long challenge for my taste b…
2003  I liked that I was able to occupy my reading a…
2004  I think it's good not only to research other t…

     feedback_challenge5_sentiment feedback_challenge5_sentiment_confidence
0                         Positive                                 0.892941
1                          Neutral                                 0.495489
2                         Positive                                 0.937368
3                          Neutral                                 0.559524
4                          Neutral                                 0.706039
…                            …                                      …
2000                       Neutral                                 0.495489
2001                       Neutral                                 0.495489
2002                      Positive                                 0.639073
2003                      Positive                                  0.97864
2004                      Positive                                    0.858

[2005 rows x 17 columns]
```

[131]:
```
df2.to_excel('data/student_feedback_sentiment_analysys.xlsx',␣
↪sheet_name='feedback_sentiment')
```

[9]:
```
# Open xlsx file
df_sa = pd.read_excel('data/student_feedback_sentiment_analysis.xlsx',␣
↪sheet_name='feedback_sentiment')
```

[10]:
```
df_sa.head()
```

[10]:
```
   student_id survey_month  survey_year         before_lifepurpose
0  AL02674306    September         2021  Neither agree or disagree  \
1  AL02684742    September         2021  Neither agree or disagree
2  AL02685313         June         2022              Totally agree
3  AL02685686    September         2021           Partially agree
4  AL02685806    September         2021  Neither agree or disagree

  before_personalresources   before_motivation          after_lifepurpose
0          Partially agree       Totally agree  Neither agree or disagree  \
1       Partially disagree  Partially disagree         Partially disagree
2            Totally agree       Totally agree              Totally agree
3          Partially agree     Partially agree              Totally agree
4       Partially disagree     Partially agree            Partially agree
```

```
     after_personalresources    after_motivation        recommendation_book
0            Partially agree       Totally agree             Totally agree  \
1         Partially disagree  Partially disagree  Neither agree or disagree
2              Totally agree       Totally agree             Totally agree
3              Totally agree       Totally agree             Totally agree
4  Neither agree or disagree     Partially agree  Neither agree or disagree

   recommendation_challenge5
0              Totally agree  \
1  Neither agree or disagree
2              Totally agree
3              Totally agree
4  Neither agree or disagree

                                       feedback_book feedback_book_sentiment
0   It is a story related to situations of many p…                Positive  \
1                                      Interesting                  Positive
2                              An inspiring story.                  Positive
3  It seems like a very simple story, even for ch…                Positive
4                                       It is good                  Positive

  feedback_book_sentiment_confidence
0                           0.838734  \
1                            0.73105
2                           0.921101
3                            0.61259
4                           0.967371

                                feedback_challenge5
0   it is a simple challenge but in addition to t…  \
1                                               No
2         It helped me think in a more positive way.
3                                             none
4                I have no comment on the challenge.

  feedback_challenge5_sentiment feedback_challenge5_sentiment_confidence
0                      Positive                                 0.892941
1                       Neutral                                 0.495489
2                      Positive                                 0.937368
3                       Neutral                                 0.559524
4                       Neutral                                 0.706039
```

[11]:
```
# subset df_sa to only positive feedback_book_sentiment
df_sa_positive = df_sa[df_sa['feedback_book_sentiment'] == 'Positive']
```

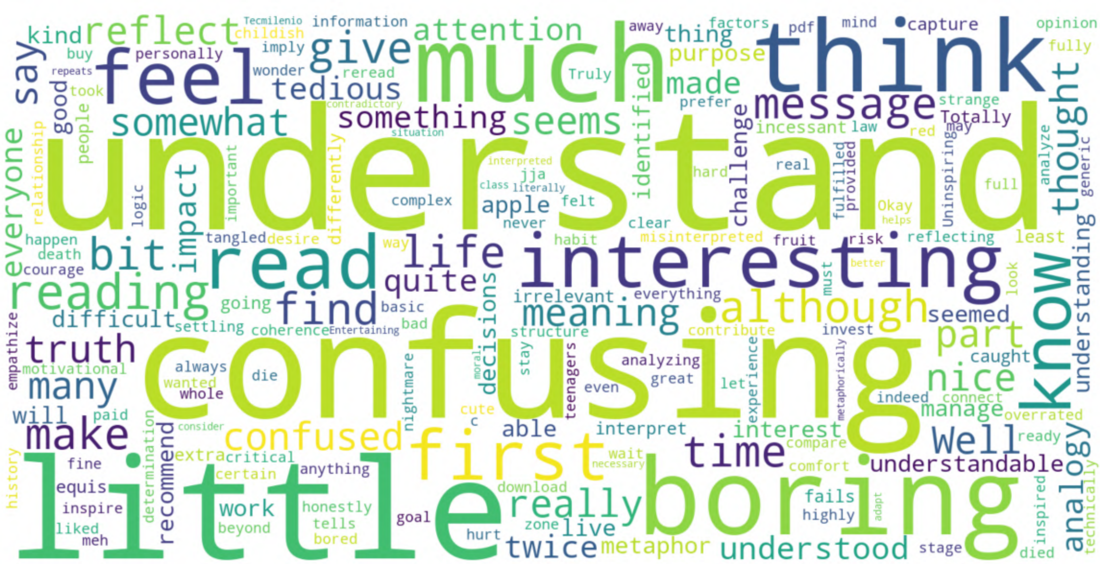
[14]:
```
from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt
```

[65]:
```
# Make a word cloud from the positive feedback_book_sentiment
# Create a list of word
text = " ".join(review for review in df_sa_positive.feedback_book.astype(str))
# print ("There are {} words in the combination of all review.".
 ↪format(len(text)))

# Create stopword list:
stopwords = set(STOPWORDS)

# Generate a word cloud image
wordcloud = WordCloud(stopwords=stopwords, background_color="white",␣
 ↪width=1500, height=750).generate(text)

# Display a big generated image:
# the matplotlib way:
plt.figure(figsize=(10, 10), dpi=1250)
plt.imshow(wordcloud, interpolation='none')
plt.axis("off")
plt.show()


# Save the image in the img folder in a bit bigger size:
wordcloud.to_file("img/wordcloud_positive_feedback_book.png")
```

[65]: <wordcloud.wordcloud.WordCloud at 0x205865855d0>

[66]:

```python
# Make a word cloud from the negative feedback_book_sentiment
# subset df_sa to only negative feedback_book_sentiment
df_sa_negative = df_sa[df_sa['feedback_book_sentiment'] == 'Negative']

# Create a list of word
text = " ".join(review for review in df_sa_negative.feedback_book.astype(str))
# print ("There are {} words in the combination of all review.".
 ↪format(len(text)))

# Create stopword list:
stopwords = set(STOPWORDS)
# append more stopwords
stopwords.update(['found', 'story'])

# Generate a word cloud image
wordcloud = WordCloud(stopwords=stopwords, background_color="white",␣
 ↪width=1500, height=750).generate(text)

# Display the generated image:
# the matplotlib way:
plt.figure(figsize=(10, 10), dpi=1250)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()

# Save the image in the img folder in a bit high resolution:
wordcloud.to_file("img/wordcloud_negative_feedback_book.png")
```



[66]: <wordcloud.wordcloud.WordCloud at 0x20589fae310>

[67]:
```
# Make a word cloud from the positive feedback_challenge5_sentiment
# subset df_sa to only positive feedback_challenge5_sentiment
df_sa_positive = df_sa[df_sa['feedback_challenge5_sentiment'] == 'Positive']

# Create a list of word
text = " ".join(review for review in df_sa_positive.feedback_challenge5.
 ↪astype(str))
# print ("There are {} words in the combination of all review.".
 ↪format(len(text)))

# Create stopword list:
stopwords = set(STOPWORDS)
# append more stopwords
stopwords.update(['challenge'])

# Generate a word cloud image
wordcloud = WordCloud(stopwords=stopwords, background_color="white",␣
 ↪width=1500, height=750).generate(text)

# Display the generated image:
# the matplotlib way:
plt.figure(figsize=(10, 10), dpi=1250)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()

# Save the image in the img folder in a bit high resolution:
wordcloud.to_file("img/wordcloud_positive_feedback_challenge5.png")
```

[67]: <wordcloud.wordcloud.WordCloud at 0x2059f1821d0>

[68]:
```
# Make a word cloud from the negative feedback_challenge5_sentiment
# subset df_sa to only negative feedback_challenge5_sentiment
df_sa_negative = df_sa[df_sa['feedback_challenge5_sentiment'] == 'Negative']

# Create a list of word
text = " ".join(review for review in df_sa_negative.feedback_challenge5.
 ↪astype(str))
# print ("There are {} words in the combination of all review.".
 ↪format(len(text)))

# Create stopword list:
stopwords = set(STOPWORDS)
# # append more stopwords
# stopwords.update(['challenge'])

# Generate a word cloud image
wordcloud = WordCloud(stopwords=stopwords, background_color="white",␣
 ↪width=1500, height=750).generate(text)

# Display the generated image:
# the matplotlib way:
plt.figure(figsize=(10, 10), dpi=1250)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()

# Save the image in the img folder in a bit high resolution:
wordcloud.to_file("img/wordcloud_negative_feedback_challenge5.png")
```

[68]: <wordcloud.wordcloud.WordCloud at 0x2058547e8d0>

[69]:
```
# Make a word cloud from the neutral feedback_challenge5_sentiment
# subset df_sa to only neutral feedback_challenge5_sentiment
df_sa_neutral = df_sa[df_sa['feedback_challenge5_sentiment'] == 'Neutral']

# Create a list of word
text = " ".join(review for review in df_sa_neutral.feedback_challenge5.
 ↪astype(str))
# print ("There are {} words in the combination of all review.".
 ↪format(len(text)))

# Create stopword list:
stopwords = set(STOPWORDS)
# # append more stopwords
# stopwords.update(['challenge'])

# Generate a word cloud image
wordcloud = WordCloud(stopwords=stopwords, background_color="white",␣
 ↪width=1500, height=750).generate(text)

# Display the generated image:
# the matplotlib way:
plt.figure(figsize=(10, 10), dpi=1250)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

```
# Save the image in the img folder in a bit high resolution:
wordcloud.to_file("img/wordcloud_neutral_feedback_challenge5.png")
```

[69]: <wordcloud.wordcloud.WordCloud at 0x2058536b390>

[ ]: