Homonyms problem in Sentiment Analysis:

- ## Introduction:

  Homonyms in the sentence can pose a challenge for natural language processing tasks such as sentiment analysis. The Homonyms in sentence in sentiment analysis means that the same word may have two different sentiments with two different meaning for the whole sentence.

  In the examples provided, “I hate the selfishness in you” and “I hate anyone who can hurt you” the word "hate" has two different sentiments for each sentence, the first one has negative sentiment, but in the second one it has positive sentiment, however verb “hate” may leads to negative sentiment, but by analyzing the meaning of the sentence may be positive.

  To achieve great classification on homonyms problem it’s necessary to choose good pretrained model and retrain it into dataset which have different sentences with different meanings using the same word. Then calculate the accuracy for both positive and negative probability and select the highest value as sentiment label.

- ## Data description:

  Dataset depends on the domain.

- ## Baseline experiments:

  The baseline experiment for this task is getting good classifier and tokenizer to overcome the problem of understanding the respective sentiment of the sentence. With search the Bert large model is the most accurate. Load pretrained model to tokenize text and make classification for sentiment.

Fig.1. Load pretrained model to tokenize text and make classification for sentiment.

```
I hate the selfishness in you
Positive probability: 32.8792929649353%
Negative probability: 67.1207070350647%
```

```
I hate any one who can hurt you
Positive probability: 99.50632452964783%
Negative probability: 0.49367458559572697%
```

```
Seeing you hurt breaks my heart into pieces
Positive probability: 84.628164768219%
Negative probability: 15.371832251548767%
```

```
It's important to remember that healing from the pain of being hurt is a process
Positive probability: 98.64650964736938%
Negative probability: 1.3534879311919212%
```

Fig.2. Testing examples in the model

From these examples it can be seen that the performance of the model is good, but with some examples not, so we need to retrain the model with dataset which fits the cases that we need.

- **The questions Answer:**

1) **What was the biggest challenge you faced when carrying out this project?**

The biggest challenge is to select the best pretrained model to Differentiates between the homonyms problem which one is negative and which one is positive through understanding the meaning of the context and the second problem were happened if the pretrained model fit with domain of data well and in another domain make wrong prediction here the problem can be solved by retrain the model with data in the domain that model doesn't fit well with it.

2) **What do you think you have learned from the project?**

There are many models can be used to classify sentiment problem but not fit well with test data because they fit with training data only and using deep learning like LSTM may have great accuracy because model prone to overfit solving this problem in using pretrained model to overcome overfitting because they are trained on large datasets , These models can then be used as a starting point for training new models or for performing specific tasks without having to start from scratch. And models can be fine-tuned on smaller datasets to solve specific problems in related domains.

- **Conclusion:**

  Homonyms problem in text analysis is a significant challenge for natural language processing tasks such as sentiment analysis. Homonyms in sentence in sentiment analysis are words that have different sentiments with different meanings for whole the sentence, which can lead to confusion in text analysis models. This can result in inaccurate sentiment analysis, which is a critical problem for businesses and researchers who rely on accurate insights from text data. To solve the homonyms problem, natural language processing models must be trained to distinguish between the different meanings of homonyms and its respective sentiments. This requires careful data preparation and advanced machine learning techniques, by trying several pretrained models, Bert large model was the most accurate, its performance is good, but with some examples not, so we need to retrain the model with dataset which fits the cases that we need .Overall, solving the homonyms problem in text analysis is essential for improving the accuracy and reliability of sentiment analysis and other natural language processing tasks.

- **Resources:**
  - https://towardsdatascience.com/fine-tuning-bert-for-text-classification-54e7df642894
  - https://blog.smart-tribune.com/en/semantic-analysis