

# Sentiment Analysis Using RoBERTa

---

Aarya Vasantlal, Sahil Gandhi, Ishayu Ray

# What is RoBERTa?



RoBERTa is mainly used for natural language processing tasks which includes translation, text-classification and answering questions



Processes input sequences and produces conceptualized outputs with the use of self-attention

Self-attention relationship between words and model

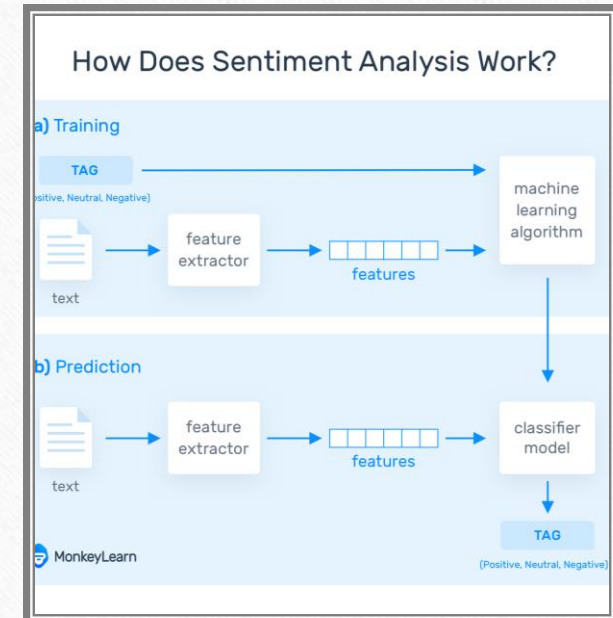


Model is based of BERT architecture and is optimized



# Problems/Tasks

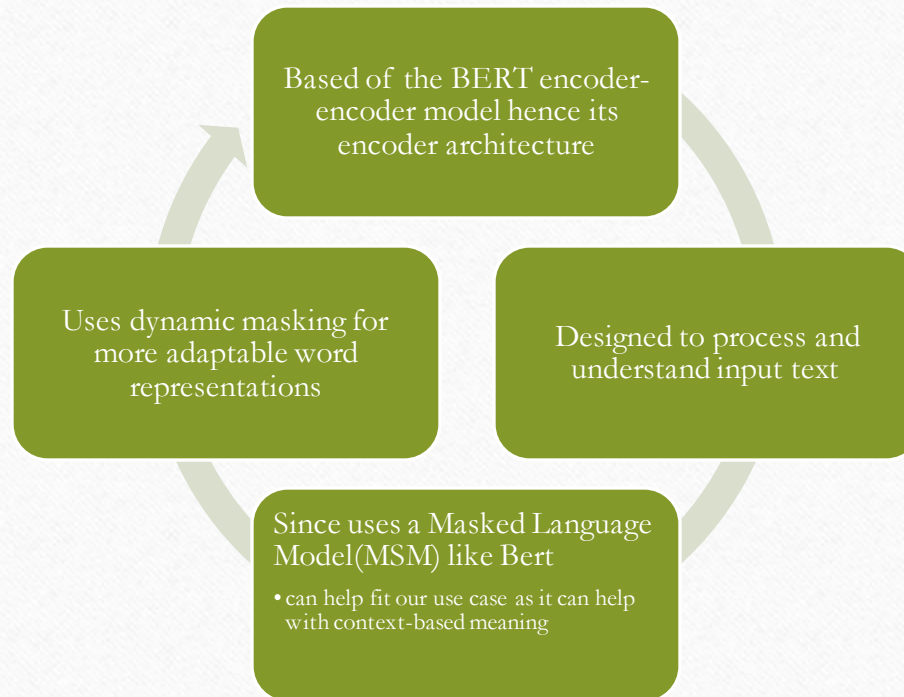
- Using the encoder Roberta, our task is to do a sentiment analysis on movie reviews using a dataset from Kaggle
- This task can also be done with Amazon product reviews, twitter, YouTube and Instagram comments are some examples
- Our task within Sentiment analysis is to find movie review comments and see if they are negative, neutral or positive comments



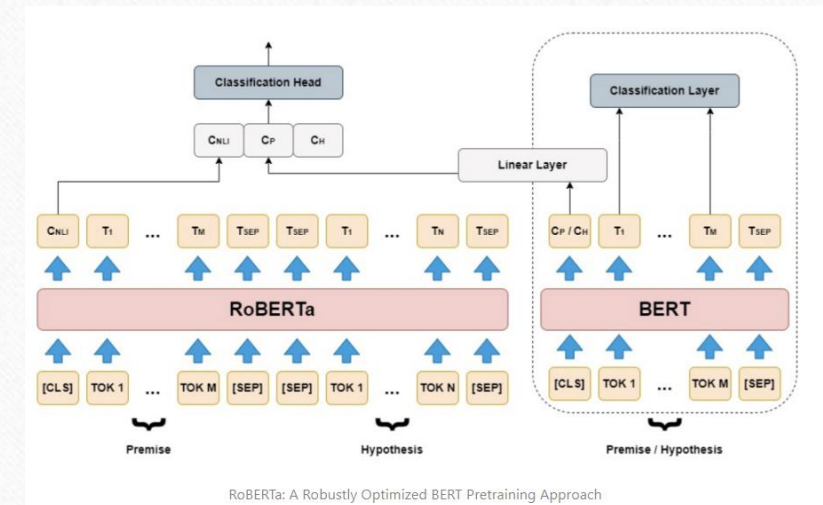
# Pretrained model/datasets

- For the pretraining of the model we will use 80% of data within our selected database from Kaggle to help the model predict what words are considered as positive, negative, or neutral. Including mixed sentiment
- To help the model better understand what words would fit this criteria we would have the model train on just singular words in a sentence and parts of a sentence to help the monitor understand what words would fit into the specified categories
- For the dataset we are using Movie Reviews from the website Rotten Tomatoes to provide us with tone-based groups of the text that can help the model train properly on the tone of the sentence.

# RoBERTa Model & Architecture



Visual Representation of the architecture





# Process & Results

---

1. Importing Python Libraries and preparing the environment
2. Preparing the Dataset & Data Loader
3. Creating the Neural Network for Fine Tuning
4. Fine Tuning the Model
5. Validating the Model
6. Saving the Trained Model Artifacts for Inference

<https://colab.research.google.com/drive/1tCO9GoRfhYnahfy4B65GG2OaPNJu7pDx#scrollTo=8OhO4xlwqExT>

# Sources

---

- [Link to the model RoBERTa](#)
- [Dataset Used from Kaggle](#)