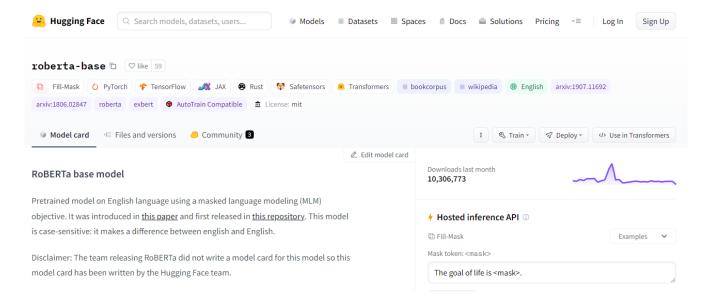


## Objective

In this project, I scraped 40000 tweets from Twitter about the "UCL Final" between Liverpool and Real Madrid, the tweets are categorized into Neutral, Positive, and Negative. Because limitation of time and resource, only 10000 tweet proceed.

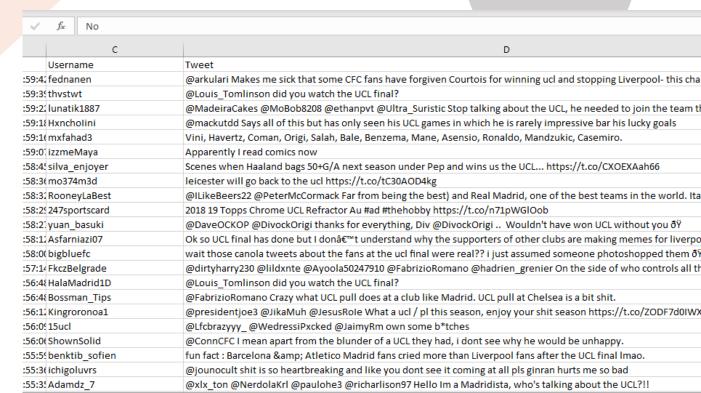
# Objective

The roBERTa-base model is used to do the Twitter sentiment analysis.



### **Dataset**

### 40000 tweets contain "UCL Final" word



#### Clean Tweet From Unwanted String

```
df.Tweet = df.Tweet.replace(r'\n',' ', regex=True)
def preprocess(text):
    new text = []
    for t in text.strip().split(" "):
        t = '@user' if t.startswith('@') and len(t) > 1 else t
        t = 'http' if t.startswith('http') else t
        new text.append(t)
    return " ".join(new text)
df['Tweet'] = df['Tweet'].map(lambda x: preprocess(x))
df.Tweet.head()
     Quser Makes me sick that some CFC fans have fo...
                   @user did you watch the UCL final?
    Quser Quser Quser Stop talking about the...
    Quser Says all of this but has only seen his U...
    Vini, Havertz, Coman, Origi, Salah, Bale, Benz...
Name: Tweet, dtype: object
```

#### Load Model and Tokenizer

```
roberta = f"cardiffnlp/twitter-roberta-base-sentiment"
model = AutoModelForSequenceClassification.from_pretrained(roberta)
tokenizer = AutoTokenizer.from_pretrained(roberta)
model.save_pretrained(roberta)
tokenizer.save_pretrained(roberta)
labels = ['Negative', 'Neutral', 'Positive']
```

#### Sentiment Analysis

```
def sentiment(tweet):
    encoded_tweet = tokenizer(preprocess(tweet), return_tensors='pt')
    output = model(encoded_tweet['input_ids'], encoded_tweet['attention_mask'])
    scores = output[0][0].detach().numpy()
    scores = softmax(scores)
    max_value = max(scores)
    list_scores = list(scores)
    index = list_scores.index(max_value)
    # print(labels[index], max_value)
    return pd.Series([max_value, labels[index]])
```

```
df[['Score', 'Label']] = df['Tweet'].apply(lambda x: sentiment(x))
```

## Result of model

	No	Date	Username	Tweet	Score	Label
0	1	2022-05-30 23:59:42+00:00	fednanen	@user Makes me sick that some CFC fans have fo	0.859798	Negative
1	2	2022-05-30 23:59:39+00:00	thvstwt	@user did you watch the UCL final?	0.880020	Neutral
2	3	2022-05-30 23:59:22+00:00	lunatik1887	@user @user @user Stop talking about the	0.625723	Neutral
3	4	2022-05-30 23:59:18+00:00	Hxncholini	@user Says all of this but has only seen his U	0.506189	Neutral
4	5	2022-05-30 23:59:16+00:00	mxfahad3	Vini, Havertz, Coman, Origi, Salah, Bale, Benz	0.716288	Neutral

### Result of model

