

# 1. Project Explanation

## 1.1 Project Idea

In this project, I used a pre-trained model from Hugging Face to do sentiment analysis. The idea was to see how well the model could determine if Yelp reviews are positive or negative using a small part of the dataset.

## 1.2 Goal

The goal was to test the pre-trained sentiment analysis model on a small subset of Yelp reviews and see how it performs without any additional fine-tuning firstly.

### ***What is Yelp ?***

*Yelp is an online platform that allows users to review and rate local businesses. Yelp was founded in 2004 in San Francisco by former PayPal employees. Yelp is a social site where users can find, contact, and share information about businesses in various categories, such as restaurants, hotels, dentists, and gyms. Yelp reviews can influence the reputation and popularity of businesses, as well as the decisions of potential customers.*

*Source: Bing*

# 2. Research and Setup

## 2.1 Research

I looked into Hugging Face Transformers, which provides many pre-trained models for different tasks. I chose the ***distilbert/distilbert-base-uncased-finetuned-sst-2-english*** model because it's the know as more beginner friendly and simple for sentiment analysis and is efficient.

## 2.2 Model Selection

The model I used is DistilBERT. It's a smaller and faster version of BERT, which has been trained to understand sentiments in text. It's useful for quickly getting results on sentiment analysis.

### 3. Implementation

#### 3.1 Loading and Preprocessing the Dataset

```
!pip install transformers
```

```
!pip install datasets
```

```
from datasets import load_dataset
```

```
# To Load the Yelp Reviews Dataset
```

```
dataset = load_dataset("yelp_review_full")
```

```
# Select a subset of 100 samples from the dataset
```

```
train_subset = dataset["train"].select(range(100))
```

#### 3.2 Loading and Testing the Model

```
from transformers import pipeline
```

```
# To load the sentiment analysis pipeline
```

```
classifier = pipeline("sentiment-analysis", model="distilbert/distilbert-base-uncased-finetuned-sst-2-english")
```

```
# To test the model with multiple reviews from the subset
```

```
results = []
```

```
for i in range(10): # Testing the first 10 reviews
```

```
    review = train_subset[i]["text"]
```

```
    sentiment = classifier(review)
```

```
    results.append((review, sentiment))
```

```
# To show the results
```

```
for i, (review, sentiment) in enumerate(results):
```

```
    print(f"Review {i+1}: {review}")
```

```
    print(f"Sentiment: {sentiment}")
```

```
    print("-" * 40)
```

### 3.3 Detailed Explanation

- **Dataset Loading:** I loaded the Yelp Reviews dataset and picked the first 10 of 100 reviews to work with.
- **Model Pipeline:** I used the pipeline function from Hugging Face to set up the sentiment analysis with the pre-trained DistilBERT model.
- **Testing:** I ran 10 sample reviews through the model to check their sentiments.

## 4. Results

### 4.1 Sample Output

Samples review from the subset:

**Review 1:** dr. goldberg offers everything i look for in a general practitioner. he's nice and easy to talk to without being patronizing; he's always on time in seeing his patients; he's affiliated with a top-notch hospital (nyu) which my parents have explained to me is very important in case something happens and you need surgery; and you can get referrals to see specialists without having to see him first. really, what more do you need? i'm sitting here trying to think of any complaints i have about him, but i'm really drawing a blank.

**Sentiment:** [{"label": "NEGATIVE", "score": 0.9836545586585999}]

-----

**Review 2:** Unfortunately, the frustration of being Dr. Goldberg's patient is a repeat of the experience I've had with so many other doctors in NYC -- good doctor, terrible staff. It seems that his staff simply never answers the phone. It usually takes 2 hours of repeated calling to get an answer. Who has time for that or wants to deal with it? I have run into this problem with many other doctors and I just don't get it. You have office workers, you have patients with medical needs, why isn't anyone answering the phone? It's incomprehensible and not work the aggravation. It's with regret that I feel that I have to give Dr. Goldberg 2 stars.

**Sentiment:** [{"label": "NEGATIVE", "score": 0.99965500831604}]

-----

**Review 3:** Been going to Dr. Goldberg for over 10 years. I think I was one of his 1st patients when he started at MHMG. He's been great over the years and is really all about the big picture. It is because of him, not my now former gyn Dr. Markoff, that I found out I have fibroids. He explores all options with you and is very patient and understanding. He doesn't judge and asks all the right questions. Very thorough and wants to be kept in the loop on every aspect of your medical health and your life.

**Sentiment:** [{"label": "POSITIVE", "score": 0.9996114373207092}]

-----

**Review 4:** Got a letter in the mail last week that said Dr. Goldberg is moving to Arizona to take a new position there in June. He will be missed very much. \n\nI think finding a new doctor in NYC that you actually like might almost be as awful as trying to find a date!

**Sentiment:** [{"label": "NEGATIVE", "score": 0.9997114539146423}]

-----

**Review 5:** I don't know what Dr. Goldberg was like before moving to Arizona, but let me tell you, STAY AWAY from this doctor and this office. I was going to Dr. Johnson before he left and Goldberg took over when Johnson left. He is not a caring doctor. He is only interested in the co-pay and having you come in for medication refills every month. He will not give refills and could less about patients's financial situations. Trying to get your 90 days mail away pharmacy prescriptions through this guy is a joke. And to make matters even worse, his office staff is incompetent. 90% of the time when you call the office, they'll put you through to a voice mail, that NO ONE ever answers or returns your call. Both my adult children and husband have decided to leave this practice after experiencing such frustration. The entire office has an attitude like they are doing you a favor. Give me a break! Stay away from this doc and the practice. You deserve better and they will not be there when you really need them. I have never felt compelled to write a bad review about anyone until I met this pathetic excuse for a doctor who is all about the money.

**Sentiment:** [{"label": "NEGATIVE", "score": 0.9990625977516174}]

-----

**Review 6:** Top notch doctor in a top notch practice. Can't say I am surprised when I was referred to him by another doctor who I think is wonderful and

because he went to one of the best medical schools in the country. \nIt is really easy to get an appointment. There is minimal wait to be seen and his bedside manner is great.

**Sentiment:** [{"label": "POSITIVE", "score": 0.9997119307518005}]

-----

**Review 7:** Dr. Eric Goldberg is a fantastic doctor who has correctly diagnosed every issue that my wife and I have had. Unlike many of my past doctors, Dr. Goldberg is very accessible and we have been able to schedule appointments with him and his staff very quickly. We are happy to have him in the neighborhood and look forward to being his patients for many years to come.

**Sentiment:** [{"label": "POSITIVE", "score": 0.999839186668396}]

-----

**Review 8:** I'm writing this review to give you a heads up before you see this Doctor. The office staff and administration are very unprofessional. I left a message with multiple people regarding my bill, and no one ever called me back. I had to hound them to get an answer about my bill. \n\nSecond, and most important, make sure your insurance is going to cover Dr. Goldberg's visits and blood work. He recommended to me that I get a physical, and he knew I was a student because I told him. I got the physical done. Later, I found out my health insurance doesn't pay for preventative visits. I received an \$800.00 bill for the blood work. I can't pay for my bill because I'm a student and don't have any cash flow at this current time. I can't believe the Doctor wouldn't give me a heads up to make sure my insurance would cover work that wasn't necessary and was strictly preventative. The office can't do anything to help me cover the bill. In addition, the office staff said the onus is on me to make sure my insurance covers visits. Frustrating situation!

**Sentiment:** [{"label": "NEGATIVE", "score": 0.9986262321472168}]

-----

**Review 9:** Wing sauce is like water. Pretty much a lot of butter and some hot sauce (franks red hot maybe). The whole wings are good size and crispy, but for \$1 a wing the sauce could be better. The hot and extra hot are about the same flavor/heat. The fish sandwich is good and is a large portion, sides are decent.

**Sentiment:** [{'label': 'POSITIVE', 'score': 0.9995321035385132}]

-----

**Review 10:** Decent range somewhat close to the city. The mats are pretty solid; however, the grass range needs to be tended too. It's like hitting out of US Open type rough...not very amenable to practicing. Which kind of defeats the purpose of going to a golf range...Still gets 3 stars because the range is lit up at night which is excellent for those of us who are addicted to this amazing game, but are somewhat short on time (having a job kinda sucks sometimes, no?).

**Sentiment:** [{'label': 'NEGATIVE', 'score': 0.9822463989257812}]The model successfully classified the review in their categories (Positive / Negative), demonstrating its effectiveness on the subset of the dataset.

## 5. Conclusion

The project showed that the pre-trained model can accurately classify sentiments of Yelp reviews. The model worked well on the small sample I tested. Although I didn't fine-tune the model, it provided good results right away.

## 6. Future Work

As a beginner in this, it was quite challenging and fun to learn about this. In the future, I could try fine-tuning the model with more data to see if it improves performance. This would involve preparing the dataset, setting up the model for fine-tuning, and comparing the results before and after.

## 7. Submission Files

- Code: Python code files for implementing sentiment analysis.
- PDF Report: Detailed report converted to PDF format.
- README.md: Instructions on running the code and project details.
- Dataset: The subset of the Yelp Reviews dataset (yelp\_reviews\_subset).