Restaurants receive thousands of customer reviews, but star ratings alone fail to explain why ratings increase or decline. Reviews often contain rich information about food quality, service, pricing, ambience, and operational factors, yet this information remains unstructured and difficult to analyze at scale. This limits restaurants' ability to identify the drivers of customer satisfaction and dissatisfaction

Notebook 1: Model Development

This notebook develops models for sentiment prediction and aspect extraction

```
!pip install pyspark
```

```
Requirement already satisfied: pyspark in /usr/local/lib/python3.12/dist-packages (4.0.1)
Requirement already satisfied: py4j==0.10.9.9 in /usr/local/lib/python3.12/dist-packages (from pyspark) (0.10.9.9)
```

```
import pyspark
print("PySpark installed and imported successfully!")
```

```
PySpark installed and imported successfully!
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
from pyspark.sql import SparkSession

spark = SparkSession.builder \
    .appName("YelpAnalysis") \
    .config("spark.driver.memory", "8g") \
    .getOrCreate()
```

Loading the data

```
business_path = "/content/drive/MyDrive/yelp_dataset/yelp_academic_dataset_business.json"

business_df = spark.read.json(business_path)
business_df.printSchema()
```

```
root
 |-- address: string (nullable = true)
 |-- attributes: struct (nullable = true)
 |    |-- AcceptsInsurance: string (nullable = true)
 |    |-- AgesAllowed: string (nullable = true)
 |    |-- Alcohol: string (nullable = true)
 |    |-- Ambience: string (nullable = true)
 |    |-- BYOB: string (nullable = true)
 |    |-- BYOBCorkage: string (nullable = true)
 |    |-- BestNights: string (nullable = true)
 |    |-- BikeParking: string (nullable = true)
 |    |-- BusinessAcceptsBitcoin: string (nullable = true)
 |    |-- BusinessAcceptsCreditCards: string (nullable = true)
 |    |-- BusinessParking: string (nullable = true)
 |    |-- ByAppointmentOnly: string (nullable = true)
 |    |-- Caters: string (nullable = true)
 |    |-- CoatCheck: string (nullable = true)
 |    |-- Corkage: string (nullable = true)
 |    |-- DietaryRestrictions: string (nullable = true)
 |    |-- DogsAllowed: string (nullable = true)
 |    |-- DriveThru: string (nullable = true)
 |    |-- GoodForDancing: string (nullable = true)
 |    |-- GoodForKids: string (nullable = true)
 |    |-- GoodForMeal: string (nullable = true)
 |    |-- HairSpecializesIn: string (nullable = true)
 |    |-- HappyHour: string (nullable = true)
 |    |-- HasTV: string (nullable = true)
 |    |-- Music: string (nullable = true)
 |    |-- NoiseLevel: string (nullable = true)
 |    |-- Open24Hours: string (nullable = true)
 |    |-- OutdoorSeating: string (nullable = true)
 |    |-- RestaurantsAttire: string (nullable = true)
 |    |-- RestaurantsCounterService: string (nullable = true)
 |    |-- RestaurantsDelivery: string (nullable = true)
 |    |-- RestaurantsGoodForGroups: string (nullable = true)
 |    |-- RestaurantsPriceRange2: string (nullable = true)
 |    |-- RestaurantsReservations: string (nullable = true)
 |    |-- RestaurantsTableService: string (nullable = true)
```

```
    |    |-- RestaurantsTakeOut: string (nullable = true)
    |    |-- Smoking: string (nullable = true)
    |    |-- WheelchairAccessible: string (nullable = true)
    |    |-- WiFi: string (nullable = true)
    |-- business_id: string (nullable = true)
    |-- categories: string (nullable = true)
    |-- city: string (nullable = true)
    |-- hours: struct (nullable = true)
    |    |-- Friday: string (nullable = true)
    |    |-- Monday: string (nullable = true)
    |    |-- Saturday: string (nullable = true)
    |    |-- Sunday: string (nullable = true)
    |    |-- Thursday: string (nullable = true)
    |    |-- Tuesday: string (nullable = true)
    |    |-- Wednesday: string (nullable = true)
    |-- is_open: long (nullable = true)
    |-- latitude: double (nullable = true)
    |-- longitude: double (nullable = true)
    |-- name: string (nullable = true)
```

```python
review_path = "/content/drive/MyDrive/yelp_dataset/yelp_academic_dataset_review.json"

reviews_df = spark.read.json(review_path)
reviews_df.printSchema()
```

```
root
 |-- business_id: string (nullable = true)
 |-- cool: long (nullable = true)
 |-- date: string (nullable = true)
 |-- funny: long (nullable = true)
 |-- review_id: string (nullable = true)
 |-- stars: double (nullable = true)
 |-- text: string (nullable = true)
 |-- useful: long (nullable = true)
 |-- user_id: string (nullable = true)
```

Cleand and Filter Data

```python
#mandatory
from pyspark.sql.functions import col

business_df_clean = business_df.filter(
    col("business_id").isNotNull() &
    col("categories").isNotNull()
)
```

```python
#Only need restaurant data
restaurants_df = business_df_clean.filter(
    col("categories").contains("Restaurants")
)
```

```python
#mandatory
reviews_df_clean = reviews_df.filter(
    col("review_id").isNotNull() &
    col("business_id").isNotNull() &
    col("text").isNotNull() &
    col("stars").isNotNull()
)
```

```python
#Sample Review
reviews_sample = reviews_df_clean.sample(fraction=0.05, seed=42)
```

```python
#Only restaurent reviews are required
restaurant_reviews_df = reviews_sample.join(
    restaurants_df.select("business_id"),
    on="business_id",
    how="inner"
)
```

Select Useful Fields for Sentiment Predeiction

```python
sentiment_df = restaurant_reviews_df.select("review_id", "business_id", "text", "stars")
```

```python
from pyspark.sql.functions import col, when

sentiment_df = sentiment_df.withColumn(
    "label",
    when(col("stars") <= 2, 0)   # Negative
    .when(col("stars") == 3, 1)  # Neutral
    .otherwise(2)  # Positive
)
```

```python
!pip install transformers datasets accelerate
```

```
Requirement already satisfied: transformers in /usr/local/lib/python3.12/dist-packages (4.57.3)
Requirement already satisfied: datasets in /usr/local/lib/python3.12/dist-packages (4.0.0)
Requirement already satisfied: accelerate in /usr/local/lib/python3.12/dist-packages (1.12.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.12/dist-packages (from transformers) (3.20.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.34.0 in /usr/local/lib/python3.12/dist-packages (from transformers)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.12/dist-packages (from transformers) (2.0.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from transformers) (25.0)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.12/dist-packages (from transformers) (6.0.3)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.12/dist-packages (from transformers) (2025.11.3
Requirement already satisfied: requests in /usr/local/lib/python3.12/dist-packages (from transformers) (2.32.4)
Requirement already satisfied: tokenizers<=0.23.0,>=0.22.0 in /usr/local/lib/python3.12/dist-packages (from transformers)
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.12/dist-packages (from transformers) (0.7.0)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.12/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.12/dist-packages (from datasets) (18.1.0)
Requirement already satisfied: dill<0.3.9,>=0.3.0 in /usr/local/lib/python3.12/dist-packages (from datasets) (0.3.8)
Requirement already satisfied: pandas in /usr/local/lib/python3.12/dist-packages (from datasets) (2.2.2)
Requirement already satisfied: xxhash in /usr/local/lib/python3.12/dist-packages (from datasets) (3.6.0)
Requirement already satisfied: multiprocess<0.70.17 in /usr/local/lib/python3.12/dist-packages (from datasets) (0.70.16)
Requirement already satisfied: fsspec<=2025.3.0,>=2023.1.0 in /usr/local/lib/python3.12/dist-packages (from fsspec[http]<=
Requirement already satisfied: psutil in /usr/local/lib/python3.12/dist-packages (from accelerate) (5.9.5)
Requirement already satisfied: torch>=2.0.0 in /usr/local/lib/python3.12/dist-packages (from accelerate) (2.9.0+cu126)
Requirement already satisfied: aiohttp!=4.0.0a0,!=4.0.0a1 in /usr/local/lib/python3.12/dist-packages (from fsspec[http]<=2
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub
Requirement already satisfied: hf-xet<2.0.0,>=1.1.3 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub<1.0,>
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests->transfo
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.12/dist-packages (from requests->transformers) (3.11
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.12/dist-packages (from requests->transformers)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.12/dist-packages (from requests->transformers)
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0->accelerate) (75.2
Requirement already satisfied: sympy>=1.13.3 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0->accelerate) (1
Requirement already satisfied: networkx>=2.5.1 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0->accelerate)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0->accelerate) (3.1.6)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch>=2
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.6.80 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0
Requirement already satisfied: nvidia-cudnn-cu12==9.10.2.21 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0-
Requirement already satisfied: nvidia-cublas-cu12==12.6.4.1 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0-
Requirement already satisfied: nvidia-cufft-cu12==11.3.0.4 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0->
Requirement already satisfied: nvidia-curand-cu12==10.3.7.77 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0
Requirement already satisfied: nvidia-cusolver-cu12==11.7.1.2 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.
Requirement already satisfied: nvidia-cusparse-cu12==12.5.4.2 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.
Requirement already satisfied: nvidia-cusparselt-cu12==0.7.1 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0
Requirement already satisfied: nvidia-nccl-cu12==2.27.5 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0->acc
Requirement already satisfied: nvidia-nvshmem-cu12==3.3.20 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0->
Requirement already satisfied: nvidia-nvtx-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0->ac
Requirement already satisfied: nvidia-nvjitlink-cu12==12.6.85 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.
Requirement already satisfied: nvidia-cufile-cu12==1.11.1.6 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0-
Requirement already satisfied: triton==3.5.0 in /usr/local/lib/python3.12/dist-packages (from torch>=2.0.0->accelerate) (3
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas->datasets) (
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas->datasets) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas->datasets) (2025.3)
Requirement already satisfied: aiohappyeyeballs>=2.5.0 in /usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
Requirement already satisfied: aiosignal>=1.4.0 in /usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!=4.0
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a
```

Convert to Hugging Face Dataset

```python
from datasets import Dataset

# Keep only relevant columns
hf_dataset = Dataset.from_spark(sentiment_df.select("text", "label"))
```

```
Loading dataset shards: 100%                                   40/40 [00:00<00:00, 47.65it/s]
```

Tokenize

```python
from transformers import AutoTokenizer

tokenizer = AutoTokenizer.from_pretrained("distilbert-base-uncased")

def tokenize(batch):
    return tokenizer(
        batch["text"],
        truncation=True,
        padding="max_length",
        max_length=128
    )

hf_dataset = hf_dataset.map(tokenize, batched=True)
hf_dataset = hf_dataset.remove_columns(["text"])
hf_dataset.set_format("torch")
```

```
/usr/local/lib/python3.12/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
  warnings.warn(
```

tokenizer_config.json: 100%   48.0/48.0 [00:00<00:00, 2.51kB/s]

config.json: 100%   483/483 [00:00<00:00, 39.8kB/s]

vocab.txt: 100%   232k/232k [00:00<00:00, 7.49MB/s]

tokenizer.json: 100%   466k/466k [00:00<00:00, 7.05MB/s]

Map: 100%   235995/235995 [02:41<00:00, 2629.80 examples/s]

Train / Evaluation Split

```python
hf_dataset = hf_dataset.shuffle(seed=42)

split_index = int(0.8 * len(hf_dataset))
train_dataset = hf_dataset.select(range(split_index))
eval_dataset = hf_dataset.select(range(split_index, len(hf_dataset)))
```

Initialize DistilBERT for classification

```python
from transformers import AutoModelForSequenceClassification

model = AutoModelForSequenceClassification.from_pretrained(
    "distilbert-base-uncased",
    num_labels=3  # Negative, Neutral, Positive
)
```

model.safetensors: 100%   268M/268M [00:02<00:00, 126MB/s]

```
Some weights of DistilBertForSequenceClassification were not initialized from the model checkpoint at distilbert-base-uncase
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
```

Training Arguments

```python
from transformers import TrainingArguments

training_args = TrainingArguments(
    output_dir="./results",
    per_device_train_batch_size=16,
    per_device_eval_batch_size=16,
    num_train_epochs=3,
    eval_strategy="epoch",
    save_strategy="no",
    report_to="none"
)
```

Evaluate arguments

```python
import numpy as np
from sklearn.metrics import accuracy_score, precision_recall_fscore_support

def compute_metrics(eval_pred):
    logits, labels = eval_pred
    preds = np.argmax(logits, axis=1)
    precision, recall, f1, _ = precision_recall_fscore_support(labels, preds, average="weighted")
    acc = accuracy_score(labels, preds)
    return {
        "accuracy": acc,
        "f1": f1,
        "precision": precision,
        "recall": recall
    }
```

Initialize Trainer

```python
from transformers import Trainer

trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=train_dataset,
    eval_dataset=eval_dataset,
    compute_metrics=compute_metrics
)
```

Train Model

```python
trainer.train()
```

[35400/35400 1:49:42, Epoch 3/3]

Epoch	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
1	0.320500	0.312321	0.881544	0.871254	0.868526	0.881544
2	0.251800	0.308882	0.883959	0.882144	0.880643	0.883959
3	0.179600	0.391490	0.882201	0.880399	0.878851	0.882201

```
TrainOutput(global_step=35400, training_loss=0.266321577944998, metrics={'train_runtime': 6584.1513,
'train_samples_per_second': 86.023, 'train_steps_per_second': 5.377, 'total_flos': 1.875732075171533e+16, 'train_loss':
```

Evaluate

```python
results = trainer.evaluate()
print(results)
```

[2950/2950 02:46]

```
{'eval_loss': 0.39149010181427, 'eval_accuracy': 0.8822008940867391, 'eval_f1': 0.8803987406994876, 'eval_precision': 0.8788
```

save sentiment prediction model

```python
MODEL_DIR = "/content/drive/MyDrive/restaurant_sentiment_model"

trainer.save_model(MODEL_DIR)
tokenizer.save_pretrained(MODEL_DIR)
```

```
('/content/drive/MyDrive/restaurant_sentiment_model/tokenizer_config.json',
 '/content/drive/MyDrive/restaurant_sentiment_model/special_tokens_map.json',
 '/content/drive/MyDrive/restaurant_sentiment_model/vocab.txt',
 '/content/drive/MyDrive/restaurant_sentiment_model/added_tokens.json',
 '/content/drive/MyDrive/restaurant_sentiment_model/tokenizer.json')
```

ASPECT EXTRACTION

Get reviews + business_id

```
reviews_pdf = sentiment_df.select("business_id", "text").toPandas()[:10000]
reviews_texts = reviews_pdf["text"].tolist()
```

```
pip install -U bertopic
```

```
Collecting bertopic
  Downloading bertopic-0.17.4-py3-none-any.whl.metadata (24 kB)
Requirement already satisfied: hdbscan>=0.8.29 in /usr/local/lib/python3.12/dist-packages (from bertopic) (0.8.41)
Requirement already satisfied: umap-learn>=0.5.0 in /usr/local/lib/python3.12/dist-packages (from bertopic) (0.5.9.post2)
Requirement already satisfied: numpy>=1.20.0 in /usr/local/lib/python3.12/dist-packages (from bertopic) (2.0.2)
Requirement already satisfied: pandas>=1.1.5 in /usr/local/lib/python3.12/dist-packages (from bertopic) (2.2.2)
Requirement already satisfied: plotly>=4.7.0 in /usr/local/lib/python3.12/dist-packages (from bertopic) (5.24.1)
Requirement already satisfied: scikit-learn>=1.0 in /usr/local/lib/python3.12/dist-packages (from bertopic) (1.6.1)
Requirement already satisfied: sentence-transformers>=0.4.1 in /usr/local/lib/python3.12/dist-packages (from bertopic) (5.
Requirement already satisfied: tqdm>=4.41.1 in /usr/local/lib/python3.12/dist-packages (from bertopic) (4.67.1)
Requirement already satisfied: llvmlite>0.36.0 in /usr/local/lib/python3.12/dist-packages (from bertopic) (0.43.0)
Requirement already satisfied: scipy>=1.0 in /usr/local/lib/python3.12/dist-packages (from hdbscan>=0.8.29->bertopic) (1.1
Requirement already satisfied: joblib>=1.0 in /usr/local/lib/python3.12/dist-packages (from hdbscan>=0.8.29->bertopic) (1.
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.1.5->bert
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.1.5->bertopic) (202
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas>=1.1.5->bertopic) (2
Requirement already satisfied: tenacity>=6.2.0 in /usr/local/lib/python3.12/dist-packages (from plotly>=4.7.0->bertopic) (
Requirement already satisfied: packaging in /usr/local/lib/python3.12/dist-packages (from plotly>=4.7.0->bertopic) (25.0)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.12/dist-packages (from scikit-learn>=1.0->be
Requirement already satisfied: transformers<6.0.0,>=4.41.0 in /usr/local/lib/python3.12/dist-packages (from sentence-trans
Requirement already satisfied: torch>=1.11.0 in /usr/local/lib/python3.12/dist-packages (from sentence-transformers>=0.4.1
Requirement already satisfied: huggingface-hub>=0.20.0 in /usr/local/lib/python3.12/dist-packages (from sentence-transform
Requirement already satisfied: typing_extensions>=4.5.0 in /usr/local/lib/python3.12/dist-packages (from sentence-transfor
Requirement already satisfied: numba>=0.51.2 in /usr/local/lib/python3.12/dist-packages (from umap-learn>=0.5.0->bertopic)
Requirement already satisfied: pynndescent>=0.5 in /usr/local/lib/python3.12/dist-packages (from umap-learn>=0.5.0->bertop
Requirement already satisfied: filelock in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->sentence
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->sente
Requirement already satisfied: requests in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20.0->sentence
Requirement already satisfied: hf-xet<2.0.0,>=1.1.3 in /usr/local/lib/python3.12/dist-packages (from huggingface-hub>=0.20
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2->pandas>=1
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transfo
Requirement already satisfied: sympy>=1.13.3 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-tran
Requirement already satisfied: networkx>=2.5.1 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-tr
Requirement already satisfied: jinja2 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-transformer
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch>=1.1
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch>=1
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.6.80 in /usr/local/lib/python3.12/dist-packages (from torch>=1.1
Requirement already satisfied: nvidia-cudnn-cu12==9.10.2.21 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0
Requirement already satisfied: nvidia-cublas-cu12==12.6.4.1 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0
Requirement already satisfied: nvidia-cufft-cu12==11.3.0.4 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0-
Requirement already satisfied: nvidia-curand-cu12==10.3.7.77 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.
Requirement already satisfied: nvidia-cusolver-cu12==11.7.1.2 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11
Requirement already satisfied: nvidia-cusparse-cu12==12.5.4.2 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11
Requirement already satisfied: nvidia-cusparselt-cu12==0.7.1 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.
Requirement already satisfied: nvidia-nccl-cu12==2.27.5 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->se
Requirement already satisfied: nvidia-nvshmem-cu12==3.3.20 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0-
Requirement already satisfied: nvidia-nvtx-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->s
Requirement already satisfied: nvidia-nvjitlink-cu12==12.6.85 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11
Requirement already satisfied: nvidia-cufile-cu12==1.11.1.6 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0
Requirement already satisfied: triton==3.5.0 in /usr/local/lib/python3.12/dist-packages (from torch>=1.11.0->sentence-tran
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.12/dist-packages (from transformers<6.0.0,>=4.4
Requirement already satisfied: tokenizers<=0.23.0,>=0.22.0 in /usr/local/lib/python3.12/dist-packages (from transformers<6
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.12/dist-packages (from transformers<6.0.0,>=4.
Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.12/dist-packages (from sympy>=1.13.3->torch>=1
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from jinja2->torch>=1.11.0->sen
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.12/dist-packages (from requests->hugging
```

aspect extraction

```python
from bertopic import BERTopic
from sentence_transformers import SentenceTransformer

embedding_model = SentenceTransformer("all-MiniLM-L6-v2")

topic_model = BERTopic(
    embedding_model=embedding_model,
    verbose=True
)

topics, probs = topic_model.fit_transform(reviews_texts)

#SAVE MODEL

MODEL_PATH = "/content/drive/MyDrive/restaurant_sentiment_model/aspect"
topic_model.save(
    MODEL_PATH,
    serialization="pickle"
)
```

```
2025-12-25 09:41:39,896 - BERTopic - Embedding - Transforming documents to embeddings.
Batches: 100%                                   313/313 [00:14<00:00, 61.90it/s]
2025-12-25 09:41:54,868 - BERTopic - Embedding - Completed ✓
2025-12-25 09:41:54,869 - BERTopic - Dimensionality - Fitting the dimensionality reduction algorithm
2025-12-25 09:42:12,297 - BERTopic - Dimensionality - Completed ✓
2025-12-25 09:42:12,298 - BERTopic - Cluster - Start clustering the reduced embeddings
2025-12-25 09:42:12,728 - BERTopic - Cluster - Completed ✓
2025-12-25 09:42:12,734 - BERTopic - Representation - Fine-tuning topics using representation models.
2025-12-25 09:42:13,419 - BERTopic - Representation - Completed ✓
2025-12-25 09:42:13,947 - BERTopic - WARNING: When you use `pickle` to save/load a BERTopic model,please make sure that the
```