

# Enterpret

## ML Internship Assignment - Enterpret

### About the Task

#### Problem Statement

- Given a text and a "phrase" from it, detect the sentiment expressed towards the "phrase" in the text instance.

#### Examples

- cannot rely on both milk delivery and grocery | milk | Negative
- your customer service is terrible! | customer service | Negative
- i love notion as a tool | tool | Positive
- notion is a great site and an Iphone app. | notion | Positive
- Asked for a workspace name or billing email address | billing | Neutral

#### Motivation

- Aspects are **important words** that would matter to our customers, we want to be able to provide our customers with insights on how their customers feel about these important words.
- Model evaluation will be based on performance on only Positive and Negative classes since samples with neutral class lead to no insight for our users.

### Interview process

- 1 week to complete the assignment
- 1 technical round (assignment discussion + ML)
- 1 cultural round
- 6 months of internship - 70k INR/month
- conversion to full time based on performance

# Submission

Mail us the link to the google drive when you are done.

All artifacts have to be in this format in a folder in google drive

- absa/src - contains all the source needed for model inferencing
- absa/notebooks - experimental notebooks
- absa/models - contains all models
- absa/main.py - API code
- absa/src/evaluation.py - should load the model and test file, generate results and save the result to `absa/data/results/test.csv` with columns `text, aspect and label`
- training\_methodology - google doc about your training approach
  - literature survey
  - explain your approaches
  - final metrics of each approach
  - ablation study table
  - error analysis
  - conclusion
- deployment\_pipeline - google doc about your deployment pipeline in AWS for live inferencing (not a batch process)
  - You can choose any approach to deploy
  - API should be able to scale as needed
  - Load profile in live inferencing is bursty in nature i.e can have spikes of load
- requirements.txt - all dependencies

## Timeline

You have to submit the assignment in 7 days.

## Evaluation criteria

- Training approach (40)
  - Training approaches (30)
  - Evaluation metric (10)
- Code quality (20)
- Brevity and explanation quality of the docs
  - approach doc (20)
  - inference deployment doc (20)
- Deadline extension (-10 for an extension of deadline. Can be extended by only 2 more days)

## Dataset information

- The CSV files have 3 columns
  - Text: the text entry from which phrase is extracted
  - Aspect: The phrase from the text entry for which the sentiment needs to be calculated
  - Label: The sentiment label for the data
    - 0: Negative
    - 1: Neutral
    - 2: Positive
- The dataset has been cleaned to a certain extent. You can probe more.

## Files

[train file link](#)

[test file link](#)