

⬅ Back to **Blitz Dashboard**

💬 NLP

- 🔗 🔗 Emotion Detection Puzzle: From text input, identify positive & negative emotions.
- 🔗 [Start Solving](#)
- 🚀 [Detailed Baseline](#)
- 📁 [Explore Dataset](#)

## 🕵 What Is Emotion Detection Puzzle About??

Natural Language Processing is a field of Artificial Intelligence focusing on the interaction between computers and human languages. With the rise of virtual assistants like Amazon Alexa, Siri, and Google Home. NLP has become more mainstream. Recently, GPT-3, an advanced NLP model generated blogs similar to that of humans.

We, humans, are very intuitive at identifying emotions. For instance, if you look at these GIFs, you can easily identify which one is portraying positive emotion and which one is negative. In this puzzle, you will build a model that will identify positive & negative emotions from the text.

## 💪 What You'll Learn

In this puzzle, you will learn

1. Basics of Natual Language Preprocessing

### PARTICIPANTS



### NOTEBOOKS

See all

- 5  
❤️

**Bert Using Ktrain:- LB score:-0.844**  
By aman\_patkar Almost 2 years ago

2
- 0  
❤️

**Solution for submission 147738**  
By salim\_shaikh Almost 2 years ago

0
- 0  
❤️

**Solution for submission 147640**  
By aman\_patkar Almost 2 years ago

0
- 2  
❤️

**EDA**  
By aman\_patkar Almost 2 years ago

0
- 2  
❤️

**Emotion Detection Using Machine learning**  
By aman\_patkar Almost 2 years ago

0

- 2. Using a very popular & powerful python library called spaCy for language processing to see how we can preprocess out texts using spaCy and convert them into numbers.

**Let's get started!** 🚀

## The Task

This problem aims to teach a computer to distinguish between positive and negative emotions. You will be given sentences as input. Your model should be able to accurately label those sentences as **positive or negative and output asm0 or 1** respectively where **0 is for positive** and **1 is for negative**. For this challenge, all puzzles will contain a dataset in the English language.

## Explore Dataset

The dataset contains **43413 text samples**. It is distributed across training,

[Overview](#) [Notebooks](#) [Discussion](#) [Resources](#) [Submissions](#) [Create Submission](#)

- 2. **Validation set: 3475 samples**
- 3. **Test set: 8683 samples**

The CSV file contains two columns of text and labels.

- 1. **Text:** This is a sentence from humans about emotions on some subject. It could be their views on products, services, entertainment, etc.
- 2. **Label:** This represents whether an emotion is positive (0) or negative(1). Neutral text is counted as positive.

text	label
This Apple product seems really good, I want to buy this!	0
This PC lags a lot!	1

## Dataset Files

- **train.csv - (31255 samples)** This CSV file contains a text column as the sentence and a label column as the emotion of the sentence is positive or negative.
- **val.csv - (3475 samples)** This CSV file contains a text column as the sentence and a label column as the emotion of the sentence is positive or negative.
- **test.csv - (8683 samples)** This CSV file contains a text column as the sentence and a label column containing random emotions of the sentence is positive or negative. This file also serves the purpose of sample\_submission.csv

## Evaluation Criteria

During the evaluation, the **F1 score ( Weighted Average )** and **Accuracy Score** will be used to test the model's efficiency.

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

## Getting Started

The starter kit breaks down everything from downloading the dataset, loading the libraries, processing the data, creating, training, and testing the model.

**[Click here to access the basic starter kit.](#)** This will share in-depth instructions to

1. Download the necessary files
2. Setup the Alcrow-CLI environment that will help you make a submission directly via a notebook
3. Downloading dataset & importing libraries
4. Preprocessing the dataset
5. Creating the model
6. Setting the model
7. Training the model
8. Submitting the result
9. Uploading the results

Check out the starter kit **[here!](#)** 

**Follow the instructions on the starter kit. To make your first submission:**

1. Creating a submission directory
2. Use test.csv and fill the corresponding labels.
3. Save the test.csv in the submission directory. The name of the above file should be submission.csv.
4. Inside a submission directory, put the .ipynb notebook from which you trained the model and made inference and save it as original\_notebook.ipynb.
5. Overall, this is what your submission directory should look like

submission	
└─ submission.csv	
└─ original_notebook.ipynb	

## Hint to get started

The easiest way to solve this puzzle is to convert text to embedding and then use different sklearn classifiers to classify the embedding into the emotions.

You can check out one such approach in this **[notebook](#)**.

## Resource Circle

**Here are some key concepts on Natural Language Processing**

For this problem, you will be using the powerful NLP python library SpaCy. Install this library to perform all the necessary pre-processing. What's pre-processing, you ask? Here's a breakdown of important NLP vocabulary.

### 1. Tokenization

Simply put, it is segmenting text into sentences and words. It's the task of cutting a text into pieces called tokens. It might seem simple, like just removing spaces and punctuation, but it is more nuanced than that (for example, New York would be one token, despite the space between New and York). Read more about using Spacy to perform tokenization [here](#).

## 2. Stop Words

This process includes getting rid of common language articles, pronouns, and prepositions such as "and", "the", or "to" in English. These common words appear frequently but don't provide much value in creating an objective NLP model. This process focuses on frequent words that are not informative about the text. Refer to this link on how to use Spacy to remove [stopwords](#).

## 3. Stemming and Lemmatization

Stemming is the process of removing the prefix and suffix of words. Due to the nature of the English language, sometimes this can offset the word's meaning. But using a reliable model will account for the issue. Overall, stemming helps improve the speed of an NLP model.

Lemmatization reduces words to their dictionary form, which it requires detailed dictionaries in which the algorithm can look into and link words to their corresponding lemmas. This process also considers the context of the word and helps resolve any instances of disambiguation. Here's the Spacy [guide](#) on how to perform this.

## 4. Part-of-Speech tagging

This refers to marking up a word in a text (corpus) as corresponding to a particular part of speech based on its definition and context. An example of this is the popular school activity of identifying whether a word is a noun, pronoun, verb, adjective, adverb, etc. Find the Spacy documentation on this feature [here](#).

## 5. Named Entity Recognition

The NER process locates and classifies pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, and more. This process can help answer many real-world questions. Check out SpaCy's powerful NER library and its various categories over [here](#).

These and many other steps and tools required to make your submissions are included in the starter code-kit. [Check out the starter kit here](#).

---

## Get Help From Community

[Hop over to the Alcrowd Blitz discord server](#) to see ongoing discussions about this puzzle.

---




## Subscription Queries

This is one of the many free Blitz puzzles you can access forever. To access more puzzles from various domains from the Blitz Library and receive a special new puzzle in your inbox every 2 weeks, you can subscribe to Alcrowd Blitz [here](#).

PLATFORM

- [Challenges](#)
- [Alcrowd Research](#)
- [Organize a Challenge](#)
- [FAQ](#)

CONNECT

- [Blog](#)
- [Forums](#)
- [Contact](#)
- [GitLab](#) 
- [Twitter](#) 
- [Discord](#) 

LEGAL

- [Cookie consent](#)
- [Terms of Use](#)
- [Participation Terms](#)
- [Privacy](#)
- [Community Guidelines](#)
- [AI & Ethics](#)

© 2023 Alcrowd. All rights reserved.