

LEONARDO BARGIOTTI



AMAZON REVIEWS CLASSIFIER

BUSINESS AND PROJECT MANAGEMENT

PROJECT GOALS

Amazon Reviews Classifier is an application that exploit text mining to categorize reviews as positive or negative.



DATASET DESCRIPTION

Amazon Reviews

- **Source:** <https://www.kaggle.com/datasets/kritanjali/jain/amazon-reviews?select=train.csv>
- **Classes:** 1 (Negative) and 2 (Positive)
- **Volume:** training samples is 200,000 and testing 100,000.

REVIEWS CLASSIFICATION



PRE-PROCESSING



LEARNING PHASE



RESULTS

A decorative graphic on the left side of the slide, consisting of two parallel, wavy vertical lines. The inner line is yellow and the outer line is white, both set against a dark brown background.

PRE- PROCESSING

TEXT PRE-PROCESS

PRE-PROCESS

- Apply:
 - Lowercase
 - Expanding Contractions
- Removing:
 - Stopwords
 - Digits
 - Punctuation
 - Diacritics
 - HTML Tags
 - URLs
 - Extra Whitespaces

STEMMING

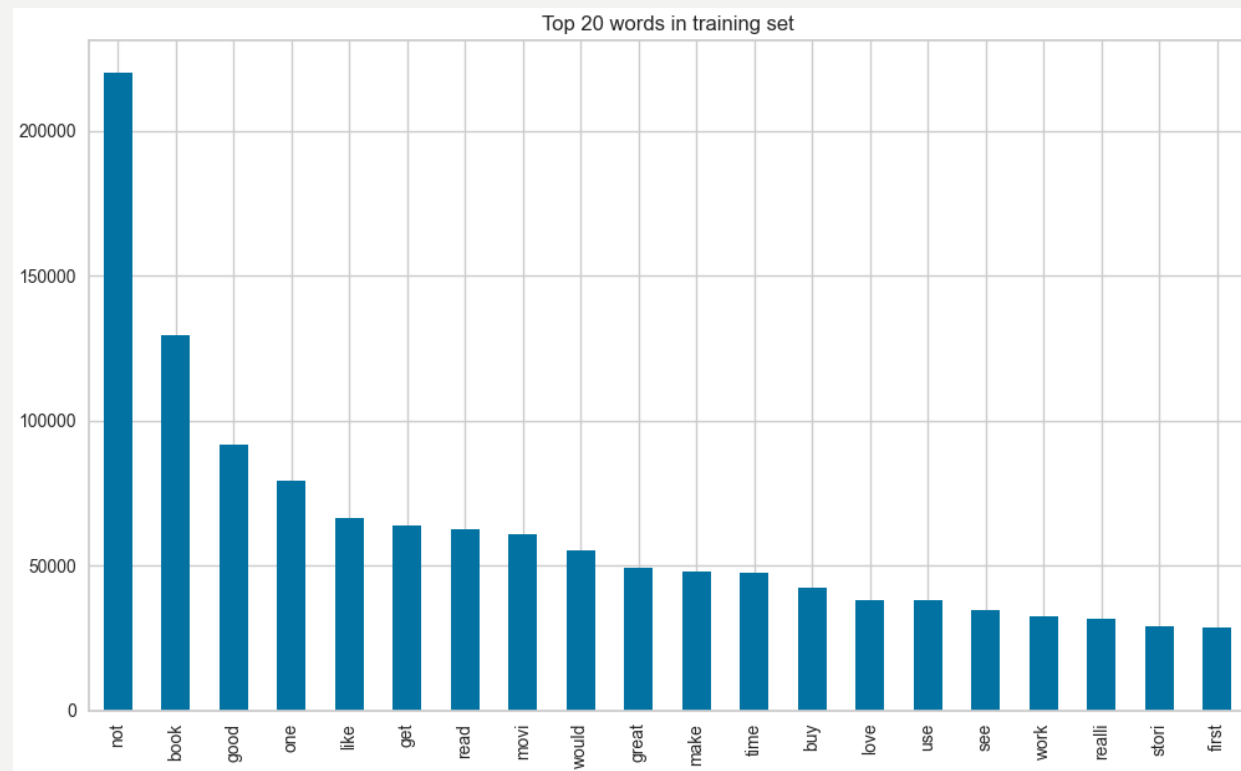
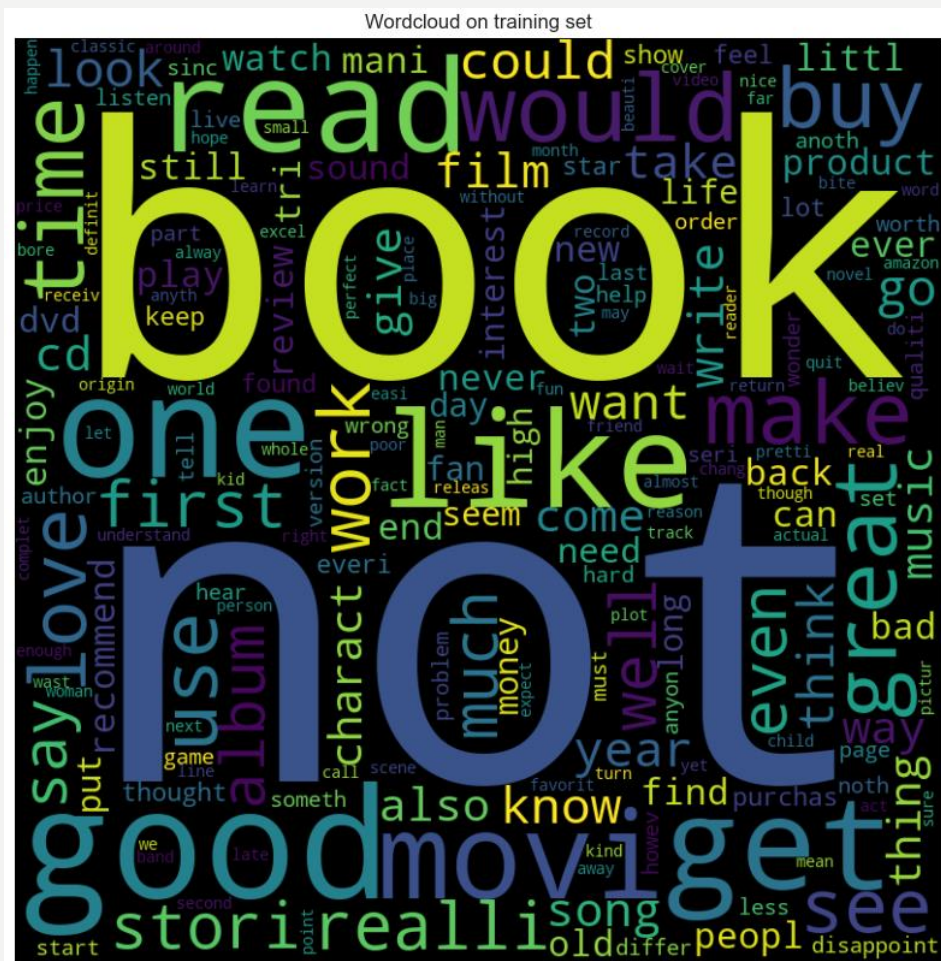
Stemming is the process of reducing words to their root form

LEMMATIZATION

The goal of lemmatization is the same as for stemming.

Lemmatization, on the other hand, is a tool that performs full morphological analysis to more accurately find the root.

PRE-PROCESS RESULTS





LEARNING PHASE

CLASSIFIERS

Traditional Classifier

Logistic Regression

Steps:

1. TfidfVectorizer

- The idea is to give more importance to terms that appear in the document but are generally less frequent

2. GridSearchCV

- Find the best Hyperparameter

Pre-Trained Classifier

DeBERTa-v3-base-mnli-fever-anli

- The base model is «DeBERTa-v3-base» of Microsoft
- Trained on multiple datasets

A decorative graphic on the left side of the slide, consisting of two parallel, wavy vertical lines. The inner line is yellow and the outer line is white, creating a stylized, organic shape.

RESULTS

RESULTS

LOGISTIC REGRESSION

	Precision	Recall	F1-Score	Support
Positive	0.89	0.89	0.89	49412
Negative	0.89	0.90	0.89	50576
Accuracy			0.89	99988
Macro Avg	0.89	0.89	0.89	99988
Weighted Avg	0.89	0.89	0.89	99988

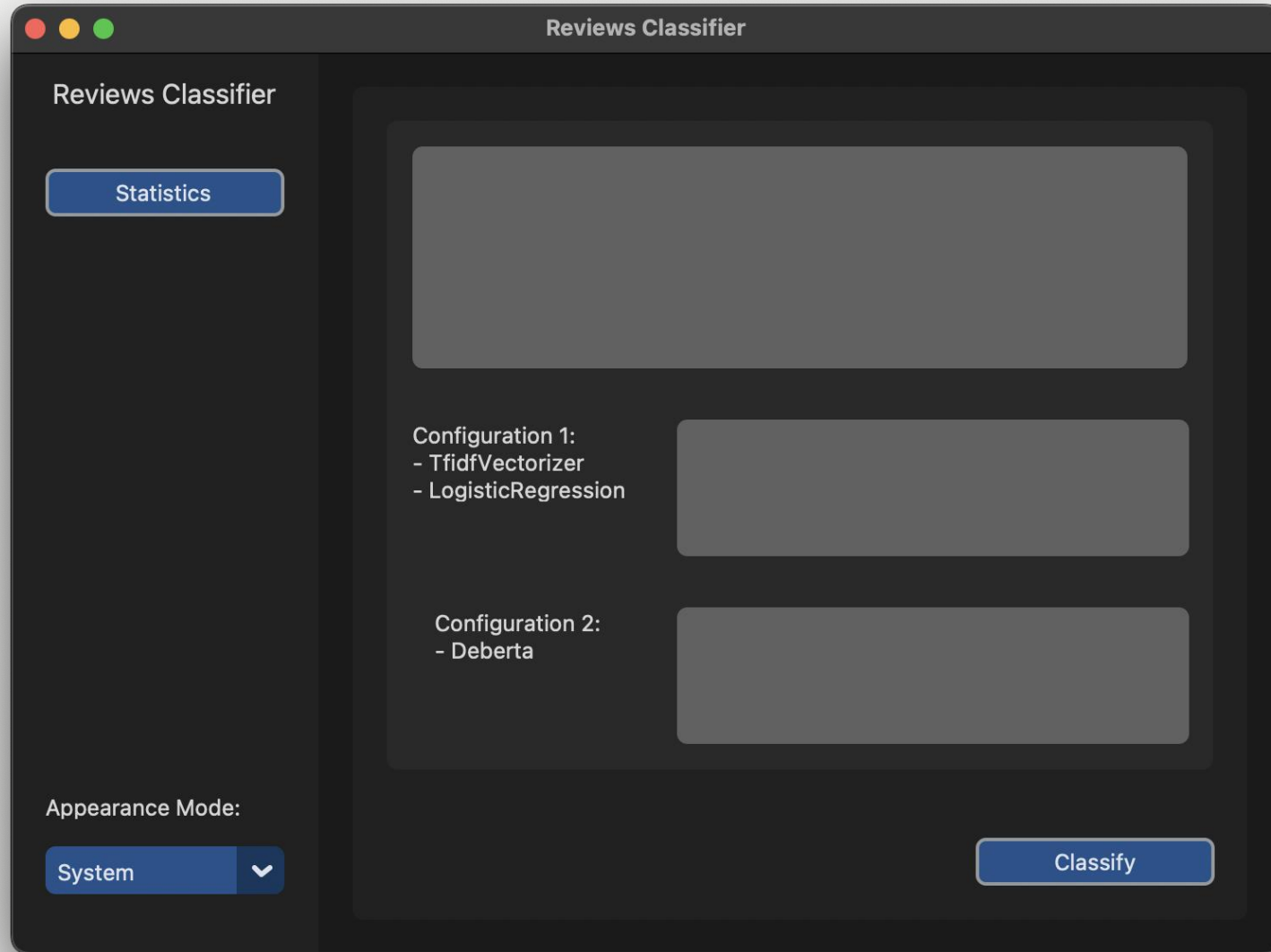
DEBERTA

	Precision	Recall	F1-Score	Support
Positive	0.87	0.89	0.88	498
Negative	0.89	0.87	0.88	502
Accuracy			0.88	1000
Macro Avg	0.88	0.88	0.88	1000
Weighted Avg	0.88	0.88	0.88	1000

A decorative wavy line in yellow and white on the left side of the image.

APPLICATION GUI

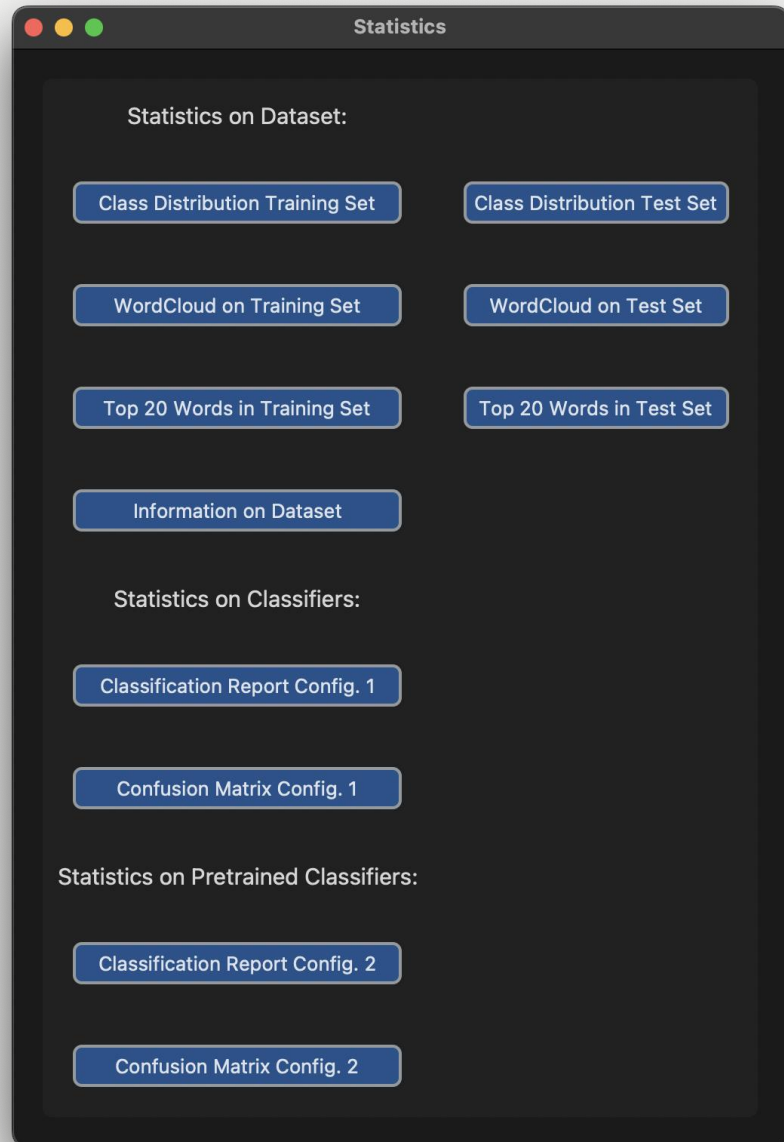
APPLICATION HOME



Features:

- 1 Text to classify
- 2 Class predicted by first classifier
- 3 Class predicted by second classifier
- 4 Button to classify text
- 5 Button to see statistics
- 6 Option menu to change appearance theme

STATISTICS WINDOW



Each button is associate to the corrispective statistic

A decorative graphic on the left side of the slide consisting of two parallel, wavy vertical lines. The inner line is bright yellow, and the outer line is white. They are set against a solid dark brown background.

EXAMPLES

POSITIVE EXAMPLE



InkSlime.

★★★★★ Such an amazing product.

Reviewed in the United States 🇺🇸 on July 4, 2023

Size: Apex 7 TKL | Style: Red – Linear & Quiet | Pattern: Keyboard | **Verified Purchase**

A very high quality product from steelseries. High quality and affordable. The keyboard Led back lighting is extremely bright and noticeable. Of course the led lighting can be dimmed. The customization of the keys are endless. The OLED screen is also very nice. Bright and highly customizable. The keys feel very nice, not too loud, very smooth, just all around amazing. The software (Steelseries GG) can be very clunky sometimes. But it is easy to use. Sometimes my custom illumination for my keys do not load up and I have to manually open up the software and select the content myself. Although a bit tedious at times but the keyboard doesn't just shut off. It defaults to one of their amazing/beautiful presets that it comes with. I highly recommend this product. Especially for people just getting into computers. I love this keyboard to death and would buy it over hundreds of times.

Reviews Classifier

Statistics

A very high quality product from steelseries. High quality and affordable. The keyboard Led back lighting is extremely bright and noticeable. Of course the led lighting can be dimmed. The customization of the keys are endless. The OLED screen is also very nice. Bright and highly customizable. The keys feel very nice, not too loud, very smooth, just all around amazing. The software (Steelseries GG) can be very clunky sometimes. But it is easy to use. Sometimes my custom illumination for my keys do not load up and I have to manually open up the software and select the content myself. Although a bit tedious at times but the keyboard doesn't just shut off. It defaults to one of their amazing/beautiful presets that it comes with. I highly recommend this product. Especially for people just getting into computers. I love this keyboard to death and would buy it over hundreds of times.

Configuration 1:
- TfIdfVectorizer
- LogisticRegression

Class predicted is: Positive
and its probability is: 96.49% in 0.053 seconds

Configuration 2:
- Deberta

Class predicted is: Positive
and its probability is: 99.66% in 0.474 seconds

Classify

Appearance Mode:

System



NEGATIVE EXAMPLE



Dave

☆☆☆☆☆ Absolutely horrendous customer service and build quality from Steelseries

Reviewed in the United States 🇺🇸 on May 21, 2021

Size: Apex 7 TKL | Style: Red – Linear & Quiet | Pattern: Keyboard | **Verified Purchase**

I've owned this keyboard for a little over a year. I've done maintenance almost 1-2 times a month to keep this keyboard as pristine as possible. Less than 6 months in, I experienced my first key chattering issue. It wasn't a big deal back then due to the fact that it was only happening on my "i" key but nonetheless, it was quite annoying to deal with. Slowly over the course of the last 3-4 months, the chattering issue spread from being only my "i" key to now infecting my k, v, h, m, d and w. Multiple inputs are registered per keystroke which rendered the keyboard basically useless to me. I created a customer support ticket and had to wait almost 3-4 weeks before I received my initial response. After back and forths for almost an additional month (keep in mind we are now 2 months into the support ticket) they finally offered me a solution. BREAK my current keyboard and send in 20 pictures to prove it was done so they can continue with an RMA, effectively rendering me USELESS for 3 weeks before my keyboard was shipped and delivered, or place a \$180 HOLD ON MY BANK ACCOUNT while they shipped out a new keyboard and I was still expected to go through the process of breaking my obviously defected keyboard before they released the hold. This is absolutely malicious and criminal activity from steelseries. I'm BEYOND angry and I would recommend everyone who sat here and read through this to NEVER PURCHASE A STEELSERIES PRODUCT. IT IS NOT worth your time and frustration.

Statistics

Configuration 1:
- TfidfVectorizer
- LogisticRegression

Class predicted is: Negative
and its probability is: 99.32% in 0.151 seconds

Configuration 2:
- Deberta

Class predicted is: Negative
and its probability is: 99.87% in 1.055 seconds

Classify

Appearance Mode:

System



CONCLUSION

TRADITIONAL CLASSIFIER

- PROS
 - Small size (40MB)
 - Less time to make a prediction (0.1s)
- CONS
 - Time to train the classifier

PRE-TRAINED CLASSIFIER

- PROS
 - Ready to use (no training phase)
- CONS
 - More time to make a prediction (1s)
 - Large size (380MB)