Sentiment Analysis

Text Analytics

Ramaseshan Ramachandran

Definition

The process of computing and classifying opinions from an unstructured text

Other names

Opinion mining, Evaluation Analysis, Appraisal Assessment, Attitude mining, Emotion extraction, subjectivity analysis, aspect extraction, affect extraction, review mining,…

Targets

- Products
  - Attributes of products,
    - Mobile phone case - design, fitment, price, durability, protection, shell type (poly carbonate, plastic), colour, weight, etc.
- Services
  - Banks, restaurants, sports centers, fitness centers, repair shops, etc.
- Organizations
  - Orgs that employ people
- Individuals
  - Fitness instructors/trainers, teachers,
- Issues
  - Political, non-political, governance, etc.
- Events
  - Music events, workshops,
- Topics

Examples

Product details

Smartphone · Single SIM · iOS · 5G · Wireless Charging · Dual Lens · With OLED Display · Facial Recognition · 2778 x 1284 · Water Resistant

The camera system receives its most significant boost yet, with next-generation technology that captures far more information, while the super-slick Pro Motion display ensures a smoother experience when surfing, gaming, or simply checking social media.

Reviews

4.6

★★★★½

11,884 reviews

5 star
4 star
3 star
2 star
1 star

Long battery life (478) | Easy to use (249) | Heavy (215) | Performs well (190) | Attractive (150) | Easy to set up (110) | Large display (110) | Quality display (86)

★★★★★ 7 December, 2021

THE BOTTOM LINE Apple's iPhone 13 Pro Max is the ultimate mobile content creation machine, with the best camera and longest battery life of any iPhone. The ultimate phone for photo and video creators! The iPhone 13 Pro Max (starting at $1,099) is the ultimate professional content creator's phone. It combines Apple's excellent camera algorithms and software support with true two-day battery life for a massive phone that's always ready to realize your dreams. While the standard iPhone 13 (starting at $799) seems to be the best choice for most people, with a terrific balance of size, power, battery life, and price, the iPhone 13 Pro Max is a terrific alternative for heavy users and artists, with its killer cameras and beautiful buttress of a battery. The iPhone 13 Pro ... More

chantal.v · Review provided by influenster.com

All reviews

Examples

Customer reviews

4.2 out of 5

2,071 global ratings

5 star	57%
4 star	23%
3 star	10%
2 star	4%
1 star	6%

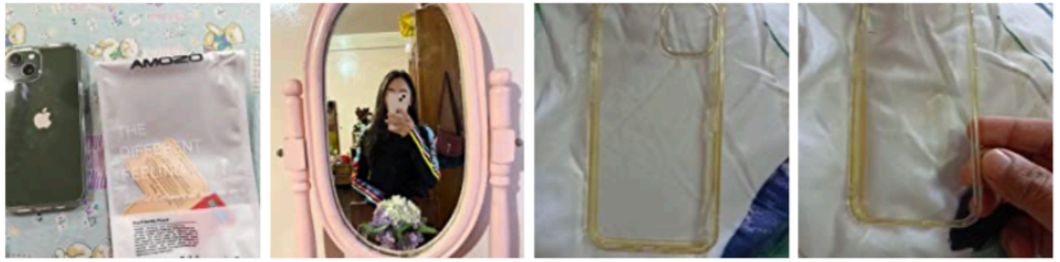
How are ratings calculated?

By feature

Value for money	4.2
Sturdiness	4.1
Durability	4.1
Sheerness	4.0

See less

Reviews with images



See all customer images

Read reviews that mention

value for money | yellow within | good quality | within a week | fits perfectly | within few days | started turning | perfect fit | fit for iphone | back cover | overall good | sturdy back | attracts a lot

Top reviews

Top reviews from India

Amazon Customer

Not sure about the Durability, But so far, so good!

Reviewed in India on 20 October 2022

Examples

Customers say

Customers like the condition and appearance of the shoes. They mention that it's value for money, the shoes are sturdy and premium. Some say that the durability is questionable and that the sole starts damaging within 2 uses. Opinions are mixed on fit, comfort, and quality.

AI-generated from the text of customer reviews

Appearance | Condition | Quality | Comfort | Fit | Performance | Durability

Customer reviews

4.1 out of 5

419 global ratings

5 star	52%
4 star	24%
3 star	14%
2 star	3%
1 star	7%

How are ratings calculated?

To calculate the overall star rating and percentage breakdown by star, we don't use a simple average. Instead, our system considers things like how recent a review is and if the reviewer bought the item on Amazon. It also analyses reviews to verify trustworthiness.

Customers say

Customers like the appearance and weight of the artificial plant. They mention it looks real and good for decoration. However, some customers are not happy that the pot is made of plastic. They are mixed on quality and size.

AI-generated from the text of customer reviews

Appearance | Weight | Quality | Size | Material | Packaging

Reviews with images

See all photo

Sample Sentences

When you don’t want to spend a whole lot of cash but want a great deal…. This is the shop to buy from

It is a very cute case

Cannot argue with the price or appearance

The jewels do fall off

It is a beautiful phone case but it is also hard to remove

Arrived broken and very flimsy

Fits perfectly but needs a little attention at the installation

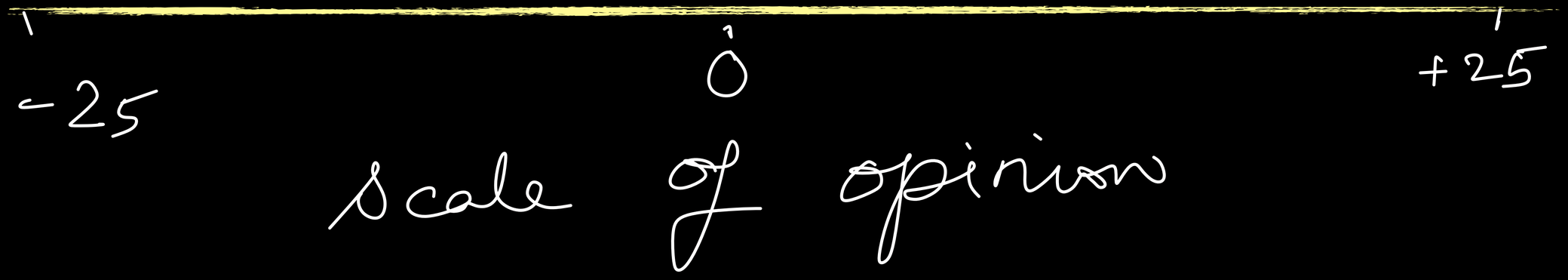
The design and colour combination makes the case simple yel elegant, and not too bold and flashy

Don’t believe that these screen protectors have glue in them

Classification Problem

Outcome

Positive, Negative, strong positive, weak positive, neutral, strong negative, weak negative, like, don't like, tone detection such as angry, friendly, formal, friendly, optimistic, etc.



Baysean Classifier for SA

Let us assume that a sentence W is represented by a a list of attributes $(w_1, w_2, w_3, \ldots, w_n)$ and let $C = \{ +, - \}$ represent the classification variable.

Using Bayes rule, the probability of the sentence W belonging to a class c is

$$p(c \mid W) = \frac{p(W \mid c)p(c)}{p(W)}$$

where $p(c \mid W), p(W \mid c)$and $p(c)$ are the posterior, likelihood and prior probabilities, respectively.

The sentence W belongs to positive sentiment only when

$$g_b(W) = \begin{cases} +, \text{if } \dfrac{p(C = + \mid W)}{p(c = - \mid W)}) \geq 1 \\ -, otherwise \end{cases}$$

where $g_b(W)$ is the ***Baysean classifier***.

Two Assumptions

1. The location and position of the word in the sentence is immaterial while computing the sentiment
    - The feature/attribute contributes to the sentiment irrespective of its position. There is no restriction wrt where word should present
    - All the attributes are independent given the value of the class variable

Naive Bayes - A Linear Classifier

Using the second assumption on conditional independence,

$$p(W \mid c) = p(w_1, w_2, w_3, \ldots, w_n \mid c)$$

$$= \prod_{j=1}^{n} p(w_j \mid c)$$

The Baysean classifier is now a Naive Bayes classifier and is of the form

$$g_b(W) = \begin{cases} +, \text{ if } \mathfrak{x} \geq 1 \\ -, \text{ otherwise} \end{cases}$$

where $\mathfrak{x} = \dfrac{p(C = +)}{p(C = -)} \prod_{j=1}^{n} \dfrac{p(w_j \mid C = +)}{p(w_j \mid C = -)}$

Multinomial Naive Bayes

The probability $p(w_j \mid c)$ can be represented in terms of the parameters $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_n)$ where x_i represents the count of the word appearing for the i^{th} instance for each sentiment $c \in C$. $\mathbf{x}$ is generated by the multinomial $(p_1, p_2, p_3, \ldots, p_n)$ where p_i is the probability that event i occurs.

We may need a smoothing parameter θ_{c_i} to avoid zero probability

$$\tilde{\theta}_{c_i} = \frac{x_{c_i} + \alpha}{x_c + \alpha \mid V \mid},$$

where x_{c_i} is the number of times the feature i appearing in the training samples for the class $c \in C$, x_c is the number of features for the class c and $\mid V \mid$ is the count of the vocabulary in the training samples.

Example

	Cat	Documents
Training	-	just plain boring
	-	entirely predictable and lacks energy
	-	no surprises and very few laughs
	+	very powerful
	+	the most fun film of the summer
Test	?	predictable with no fun

14 words

9 words

$|V| = 20$

Example

The prior $P(c)$ for the two classes is computed via Eq. 4.11 as $\frac{N_c}{N_{doc}}$:

$$P(-) = \frac{3}{5} \qquad P(+) = \frac{2}{5}$$

The word *with* doesn't occur in the training set, so we drop it completely (as mentioned above, we don't use unknown word models for naive Bayes). The likelihoods from the training set for the remaining three words "predictable", "no", and "fun", are as follows, from Eq. 4.14 (computing the probabilities for the remainder of the words in the training set is left as an exercise for the reader):

$$P(\text{"predictable"}|-) = \frac{1+1}{14+20} \qquad P(\text{"predictable"}|+) = \frac{0+1}{9+20}$$

$$P(\text{"no"}|-) = \frac{1+1}{14+20} \qquad P(\text{"no"}|+) = \frac{0+1}{9+20}$$

$$P(\text{"fun"}|-) = \frac{0+1}{14+20} \qquad P(\text{"fun"}|+) = \frac{1+1}{9+20}$$

For the test sentence S = "predictable with no fun", after removing the word 'with', the chosen class, via Eq. 4.9, is therefore computed as follows:

$\frac{3}{5} \cdot \frac{2}{34} \cdot \frac{2}{34} \cdot \frac{1}{34}$

$$P(-)P(S|-) = \frac{3}{5} \times \frac{2 \times 2 \times 1}{34^3} = 6.1 \times 10^{-5}$$ ①

$\frac{2}{5} \cdot \frac{1}{29} \cdot \frac{1}{29} \cdot \frac{2}{29}$

$$P(+)P(S|+) = \frac{2}{5} \times \frac{1 \times 1 \times 2}{29^3} = 3.2 \times 10^{-5}$$ ②

The model thus predicts the class *negative* for the test sentence.

Laplacian Smoothing is used

Here every sentence is considered as a document

Use argmax or ratio $\frac{P(w|+)P(+)}{P(w|-)P(-)}$

to find the sentiments

Words ∉ V are not considered

Is Naive Bayes a generative class?

A few observations

The test record cannot be predicted with certainty

Most effective and efficient inductive learning algorithm in machine learning

The conditional independence is not true in many real world applications.

What is the reason for its good performance?

Other Classifiers

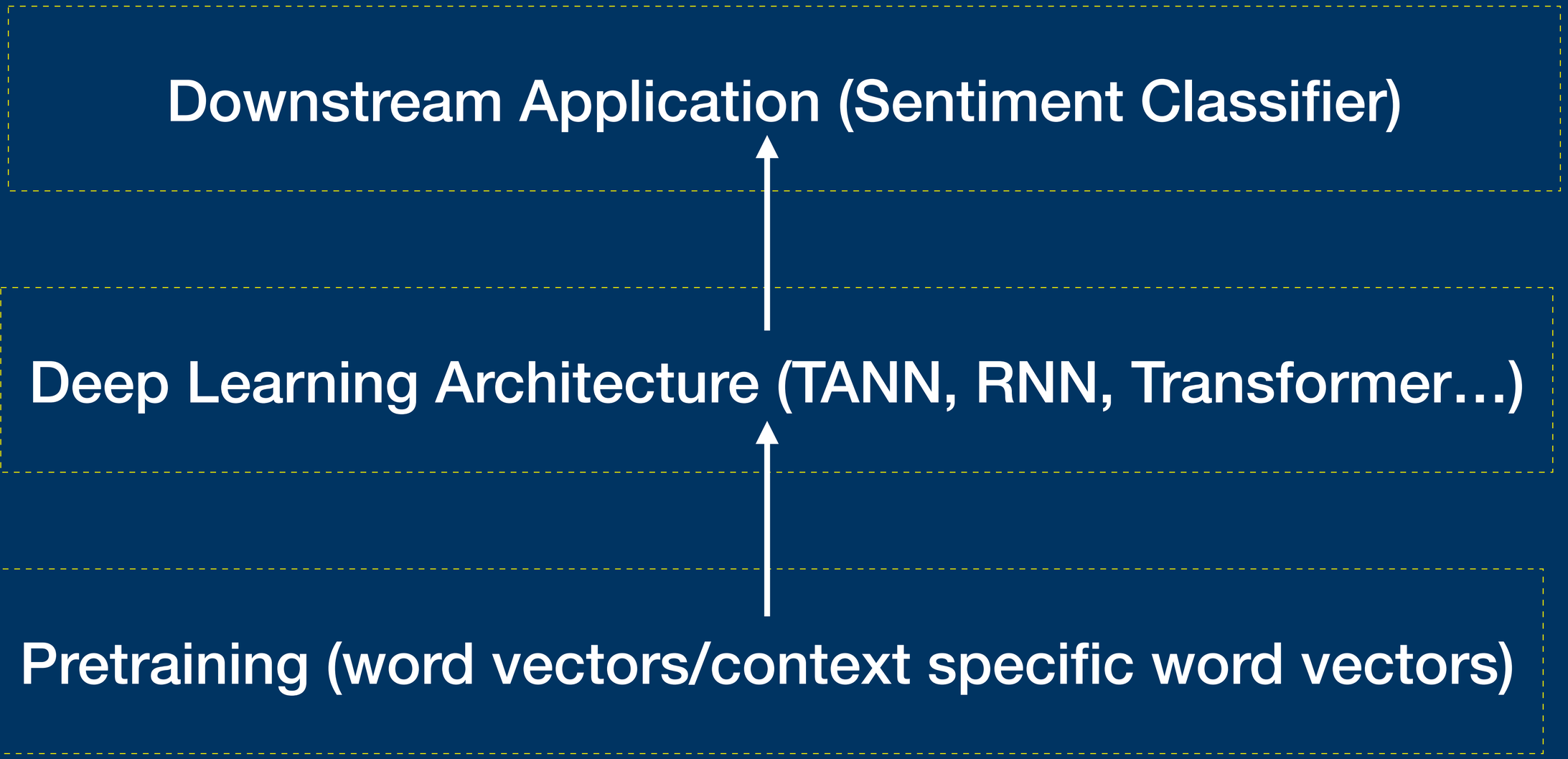
Rule based

Supervised Support Vector Machines

LSTM

Deep Learning Architecture

Deep Learning Architecture for NLP



Lexicon based approach

Lemmatize all sentences

Let $\mathfrak{L}$ represent the sentiment lexicon pairs, the word and irs polarity

$Positives = 0, Negatives = 0$

Test Sentence $S = \{w_1, w_2, w_3 \ldots w_n\}$

For w in S

    If w found in $\mathfrak{L}$ and w == positive, then

            Positives = Positives +1

    Else

            Negatives = Negatives +1

If Positives > Negatives then

    Return positive

Return negative