

Sentiment Analysis for Text using ML

Sentiment analysis using machine learning is a tool that analyses texts for polarity, from the text to either Positive or Negative or Neutral. By training machine learning tools with examples of emotions in text, machines automatically learn how to detect sentiment without human input.

Problem Statement:

Nowadays everything is getting digital, gone are those days where everyone used to go to shopping marts to buy products, now everything is just a click away. With the boom in internet companies, there is a high competition in the industry so to retain the customers, companies have the urge to analyse the feedback and evolve over time. With millions of customers, it is almost impossible to manually review the customer sentiment. That is where our problem is, we need to find the best fitting classifier which tells the sentiment of the customer based on their review of some product.

This is a text classification problem. This model predicts the sentiment of the customer from the text to either Positive or Negative or Neutral.

Data Source:

Amazon Food Reviews - <https://www.kaggle.com/snap/amazon-fine-food-reviews>

This dataset consists of ~500,000 reviews of fine foods from. The data span a period of more than 10 years. Reviews include several features like 'ProductId', 'UserId', 'Score' and 'text'.

Though there are many columns we will considering 'Text' and 'Score' columns as our input/output, respectively. Data includes:

- Reviews from Oct 1999 - Oct 2012
- Reviews from different Amazon categories.
- Number of users: 256,059
- Number of products: 74,258
- Number of reviews: 568,454
- Timespan: Oct 1999 - Oct 2012
- Number of Attributes/Columns in data: 10

Feature description:

- ProductId (Categorical Variable) - Unique identifier for the product
- UserId (Categorical Variable) - Unique identifier for the user
- ProfileName (Text) - Profile of the user
- HelpfulnessNumerator (Numerical) - Number of users who found the review helpful
- HelpfulnessDenominator (Numerical) - Number of users who indicated whether they found the review helpful or not
- Score (Ordinal) - Rating between 1 and 5
- Time (Numerical) - Timestamp for the review
- Summary (Text)- Summary of the review

- ### Sample Dataset:

A	B	C	D	E	F	G	H	I	J
ID	ProductID	Product Name	Profile Title	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
1	B0011K5G00	A3G0K074AHU8H0R	IT	1	5	1.304E+09	0	Good Quality Dog Food	I have bought several of the Vitality canned dog food products and have found
2	B00813GR64	A1D87762CEV5NK	Ill pa	0	0	1.347E+09	0	Not as Advertised	Product arrived labeled as humo Salted Peanuts...the peanuts were actually sn
3	B000L0OCHO	ABXLMWJX0AIN	Natalia Correa "Natalia Correa	1	1	4.219E+09	0	"Delight" says it all	This is a confection that has been around a few centuries. It is a light, pillow c
4	A30010A0UQ04	A395BORCF6GVXV	Karl	3	2	1.308E+09	0	Cough Medicine	If you are looking for the secret ingredient in Robitussin I believe I have found i
5	B006KZ2Z7K	A1JQ0RSL8FW61T	Michael D. Bigham "M. Wassin	0	0	5.151E+09	0	Great taffy	I Great taffy at a great price. There was a wide assortment of yummy taffy. Deli
6	B006KZ2Z7K	A1D05R81MG6U0	Twopennything	0	0	4.342E+09	0	Nice Taffy	I got a wide gift for taffy and ordered this five pound bag. The taffy was all ver
7	B006KZ2Z7K	A1310VYXR81I	David C. Sullivan	1	1	1.345E+09	0	Great! just as good as the expensive brands!	I got the saltwater taffy that had great flavors and was very soft and chewy. Each can
8	B006KZ2Z7K	A13MRGV0EN31IQ	Pamela G. Williams	0	0	1.336E+09	0	Wonderful, tasty taffy	This taffy is so good. It is very soft and chewy. The flavors are amazing. I wou
9	B000E7L2R4	A13MW9Y0T2K0BB1	R. James	1	1	5.132E+09	0	Yay Baby!	Right now I'm mostly just sprouting this so my cats can eat the grass. They love
10	A2118740VZCY14	A2118740VZCY14	Carol A. Reed	0	0	5.1351E+09	0	Healthy Dog Food	This is a very healthy dog food. Good for their digestion. Also good for small pu
11	B0001P89FE	A33HKOZ7OWQ0N4	Canadian Fan	1	3	5.108E+09	0	The Best Hot Sauce in the World	I don't know if it's the cactus or the tequila or just the unique combination of a
12	B0009MLVGG	A272584P8000	A Poeng "SparkyGoHome"	4	4	5.283E+09	0	My cats LOVE this "diet" food better than the	One of my boys needed to lose some weight and the other didn't. I just put this
13	B0009MLVGG	A3276123YH90	My Cats Are Not Fans of the New Food	1	1	1.345E+09	0	My cats are not fans of the new food	I have been happily eating this Platium for more than two years. I
14	B0009MLVGG	A18ECVX2R7HUE	Wille "rookie"	2	2	4.289E+09	0	Fresh and greasy!	I love flavor! These came securely packed... they were fresh and delicious! I love
15	B0011GV5IM	A2MUGV72TDQ47K	Lyndie "Oh Hell no"	4	5	5.268E+09	0	Strawberry Twizzlers - Yummy	The Strawberry Twizzlers are my guilty pleasure - yummy. Six pounds will be ar
16	B0011GV5IM	A1CX3C8PI8QJ	Brian A. Lee	5	5	5.262E+09	0	LOTS of twizzlers, just what you expect.	My daughter loves twizzlers and this shipment of six pounds really hit the spot
17	B0011GV5IM	A23X6FW605B280	Erie Neathery	0	0	2.348E+09	0	poor taste	I love eating them and they are good for watching TV and looking at movies! It
18	B0011GV5IM	A1F6H419V76Z60	Becky	0	0	5.345E+09	0	Love it	I've very satisfied with my Twizzler purchase. I shared these with my twizzl
19	B0011GV5IM	A29K95W6Z8BLP	Wolfee1	5	5	5.325E+09	0	GREAT SWEET CANDY!	Twizzlers, Strawberry my childhood favorite candy, made in Lancaster Pennsylv
20	B0011GV5IM	A31N70L213K2U	Greg	0	0	5.318E+09	0	Home delivered twizzlers	Candy was delivered very fast and was purchased at a reasonable price. I was i
21	B0011GV5IM	A1M00K6LPRSPV6	amomzemma	0	0	5.313E+09	0	Always fresh	My husband is a Twizzlers addict. We've bought these many times from Amaz
22	B0011GV5IM	A20V9617RGZ2H	Tammy Anderson	0	0	5.309E+09	0	TWIZZLERS	I bought these for my husband who is currently overseas. He loves these, and
23	B0011GV5IM	AR0VQL4N737A1	Charles Brown	0	0	5.305E+09	0	Delicious product!	I can remember buying this candy as a child and the quality hasn't dropped in a
24	B0011GV5IM	A1K130Z2UC5V7	Charles Brown	0	0	1.304E+09	0	Delicious product!	I love this candy. After the holidays I had to run out to buy it but still loved it
25	B0011GV5IM	A22P209N8HKE	S. C. Cavanaugh "jilly pepper"	0	0	5.295E+09	0	Please send these in Mexico!	I have lived out of the US for over 2 yrs now, and I so miss my Twizzlers!!! We
26	B0011GV5IM	A31PNR033H3P5	Deborah S. Liner "Cat Lady"	0	0	5.288E+09	0	Twizzlers - Strawberry	The Product received is as advertised -> <-/b> <-/a href="http://www.amazon.co

Rating of 4 or 5 → Positive review

Rating of 3 → Neutral review

The review text contains a detailed description of the product from the user's perspective. Moreover, it also describes the overall sentiment of the user toward the product apart from the description of the product itself. Some of the examples of the review text are shown below:

- 1) "This is great stuff. Made some really tasty banana bread. Good quality and lowest price in town."
- 2) This coffee is great because it's all organic ingredients! No pesticides to worry about plus it tastes good, and you have the healing effects of Ganoderma.
- 3) These condiments are overpriced and terrible. The classic is disgustingly sweet. The spiced tastes like a bad spicy marinara sauce from a chain restaurant.

On the other hand, the summary text represents a user's sentiment in a very confined manner. The information is conveyed in few words in this case. Some of the examples of the summary text are shown below:

- 1) "Best deal ever!"
- 2) "Waste of money"
- 3) "Great beans!!!"
- 4) "Big disappointment"

There are 5 major steps involved in the building a ML model for sentiment classification.This encapsulates the following steps:

- 1) Data Acquisition & Performing exploratory data analysis for generating the binary response variable

- 2) Performing various text pre-processing steps which are used to remove noisy terms
- 3) Modelling & tuning the data
- 4) Evaluating the Model
- 5) Deploying the model

How Does Sentiment Analysis with Machine Learning Work?

There are several techniques and complex algorithms used to command and train machines to perform sentiment analysis. There are pros and cons to each. But, used together, they can provide exceptional results.

Logistic Regression

Here we can apply the logistic regression because we know that the logistic regression works well for binary classification and for high dimensional sparse data.