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. It is expensive to check each review manually and label its sentiment. So, a better way is to rely on machine learning/deep learning models for that.

Rationale Statement:

We use logistic regression on model as logistic regression works well for binary classification and for high dimensional sparse data which provides better insights on the customer sentiments

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
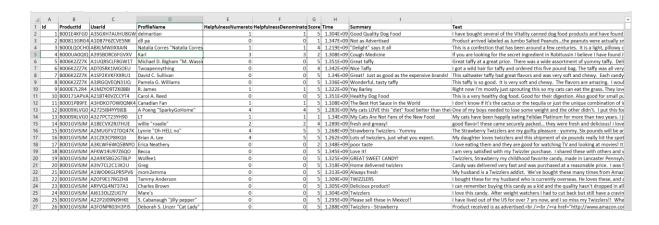- Number of reviews: 568,454
- Timespan: Oct 1999 - Oct 2012
- Number of Attributes/Columns in data: 10

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

- Text (Text) - Text of the review

Sample Dataset:

Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1.304E+09	Good Quality Dog Food	I have bought several of the Vitality canned dog food products and have found
2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1.347E+09	Not as Advertised	Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually sm
3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres	1	1	4	1.219E+09	"Delight" says it all	This is a confection that has been around a few centuries. It is a light, pillowy c
4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1.308E+09	Cough Medicine	If you are looking for the secret ingredient in Robitussin I believe I have found i
5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassi	0	0	5	1.351E+09	Great taffy	Great taffy at a great price. There was a wide assortment of yummy taffy. Deli
6	B006K2ZZ7K	ADT0SRK1MGOEU	Twoapennything	0	0	4	1.342E+09	Nice Taffy	I got a wild hair for taffy and ordered this five pound bag. The taffy was all very
7	B006K2ZZ7K	A1SP2KVKFXXRU1	David C. Sullivan	0	0	5	1.34E+09	Great! Just as good as the expensive brands!	This saltwater taffy had great flavors and was very soft and chewy. Each candy
8	B006K2ZZ7K	A3JRGQVEQN31IQ	Pamela G. Williams	0	0	5	1.336E+09	Wonderful, tasty taffy	This taffy is so good. It is very soft and chewy. The flavors are amazing. I woul
9	B000E7L2R4	A1MZYO9TZK0BBI	R. James	1	1	5	1.322E+09	Yay Barley	Right now I'm mostly just sprouting this so my cats can eat the grass. They love
10	B00171APVA	A21BT40VZCCYT4	Carol A. Reed	0	0	5	1.351E+09	Healthy Dog Food	This is a very healthy dog food. Good for their digestion. Also good for small pu
11	B0001PB9FE	A3HDKO7OW0QNK4	Canadian Fan	1	1	5	1.108E+09	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 i
12	B0009XLVG0	A2725IB4YY9JEB	A Poeng "SparkyGoHome"	4	4	5	1.283E+09	My cats LOVE this "diet" food better than thei	One of my boys needed to lose some weight and the other didn't. I put this fo
13	B0009XLVG0	A327PCT23YH90	LT	1	1	1	1.34E+09	My Cats Are Not Fans of the New Food	My cats have been happily eating Felidae Platinum for more than two years. I ju
14	B001GVISJM	A18ECVX2RJ7HUE	willie "roadie"	2	2	4	1.289E+09	fresh and greasy!	good flavor! these came securely packed... they were fresh and delicious! I love
15	B001GVISJM	A2MUGFV2TDQ47K	Lynrie "Oh HELL no"	4	5	5	1.268E+09	Strawberry Twizzlers - Yummy	The Strawberry Twizzlers are my guilty pleasure - yummy. Six pounds will be ar
16	B001GVISJM	A1CZX3CP8IKQIJ	Brian A. Lee	4	5	5	1.262E+09	Lots of twizzlers, just what you expect.	My daughter loves twizzlers and this shipment of six pounds really hit the spot
17	B001GVISJM	A3KLWF6WQ5BNYO	Erica Neathery	0	0	2	1.348E+09	poor taste	I love eating them and they are good for watching TV and looking at movies! It
18	B001GVISJM	AFKW14U97Z6QO	Becca	0	0	5	1.345E+09	Love it!	I am very satisfied with my Twizzler purchase. I shared these with others and w
19	B001GVISJM	A2A9X58G2GTBLP	Wolfee1	0	0	5	1.324E+09	GREAT SWEET CANDY!	Twizzlers, Strawberry my childhood favorite candy, made in Lancaster Pennsylv
20	B001GVISJM	A3IV7CL2C13K2U	Greg	0	0	5	1.318E+09	Home delivered twizlers	Candy was delivered very fast and was purchased at a reasonable price. I was h
21	B001GVISJM	A1WO0KGLPR5PV6	mom2emma	0	0	5	1.313E+09	Always fresh	My husband is a Twizzlers addict. We've bought these many times from Amaz
22	B001GVISJM	AZOF9E17RGZH8	Tammy Anderson	0	0	5	1.309E+09	TWIZZLERS	I bought these for my husband who is currently overseas. He loves these, and a
23	B001GVISJM	ARYVQL4N737A1	Charles Brown	0	0	5	1.305E+09	Delicious product!	I can remember buying this candy as a kid and the quality hasn't dropped in all
24	B001GVISJM	AJ613OLZZUG7V	Mare's	0	0	5	1.304E+09	Twizzlers	I love this candy. After weight watchers I had to cut back but still have a cravin
25	B001GVISJM	A22P2J09NJ9HKE	S. Cabanaugh "jilly pepper"	0	0	5	1.295E+09	Please sell these in Mexico!!	I have lived out of the US for over 7 yrs now, and I so miss my Twizzlers!! Whe
26	B001GVISJM	A3FONPR03H3PJS	Deborah S. Linzer "Cat Lady"	0	0	5	1.288E+09	Twizzlers - Strawberry	Product received is as advertised. <a href="http://www.amazon.co

We could use Score/Rating feature to determine if a review is positive or negative.

Rating of 4 or 5 → Positive review

Rating of 1 or 2 → Considered as negative one

Rating of 3 → Neutral review

Overview of the review and summary text:

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"

Methodology:

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.

We use logistic regression method for classifying our problem and measure accuracy.