Introduction to AIML - CS2011

Amazon Product Review Analysis

Team Members:

K.Sai Venkata Pranav(123ec0553)

Boda Lokesh(123ec0559)

Tanakala Harsha(123ec0549)

Kandala Nishant(123ec0569)

Dinesh Kumar Reddy(123ec0548)

Racharla Anvesh(123ei0598)

Mallepally Rishivarun(123ec0579)

Dasari Govardhan(123ec0535)

Madhav Banothu(123ei0593)

Bondugula Rishi(123ec0578)

Lokeswara Vikramaditya(123ec0544)

Praneeth Jarpla(123ei0592)

Bhukya Shireesh(123ec0568)

Tanuj Patnaik(123ec0562)

Thella Chetan Tanmai(123ec0586)

Adulla Purvi Reddy(123ec0534)

Pedagandham Srivizna(123ec0536)

V Pallavi Naidu(123ec0566)

Desam Chandini Lakshmi(123ec0845)

Nainika Challa(123ei0605)

Gajula Sai Sreeja(123ei0638)

Problem Description

The lack of efficient and automated systems for analyzing Amazon product reviews leads to several challenges:

1. **Customer Decision-making**: Buyers face difficulty identifying trustworthy and relevant information.
2. **Business Insights**: Sellers struggle to understand detailed feedback regarding specific products.
3. **Time and Effort**: Scanning reviews manually to detect patterns or trends is time-consuming and inefficient.

Goals

The project aims to develop an ML based program that can perform the following:

1. **Aspect-Based Sentiment Analysis (ABSA)**: Extract opinions on specific product features and analyze sentiments related to these features and classify reviews as positive or negative.
2. **Review Summarization**: Generate summaries of reviews to save time for customers and businesses.
3. **Spam Detection**: Identify fake or irrelevant reviews using text patterns and metadata.
4. **Trend Analysis**: Provide insights into customer satisfaction trends over time, helping businesses track product performance and customer expectations.
5. **Visualization and Reporting**: Create interactive dashboards to visualize review data, including word clouds, sentiment trends, and aspect-based insights.

Amazon hosts an extensive collection of product reviews from users across the globe, offering valuable insights into customer opinions, preferences, and experiences. Thus we are using the reviews on amazon for our project.

Plan Of Action

I. Dataset

Reviews may be collected from APIs or preexisting datasets.

Some of such datasets that we intend to use are:

https://www.kaggle.com/datasets/arhamrumi/amazon-product-reviews/data

https://kaggle.com/datasets/jillanisofttech/amazon-product-reviews

CSV File of above dataset:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Id	ProductId	UserId	ProfileNam(	Helpfulness	Helpfulness	Score	Time	Summary	Text		
2	1	B001E4KFG(	A3SGXH7AU	delmartian	1	1	5	1.304E+09	Good Qualit	I have bought several of the Vitality canne		
3	2	B00813GRG	A1D87F6ZC'	dll pa	0	0	1	1.347E+09	Not as Adve	Product arrived labeled as Jumbo Salted P(		
4	3	B000LQOCH	ABXLMWJIX	Natalia Cor|	1	1	4	1.219E+09	"Delight" sa	This is a confection that has been around ;		
5	4	B000UA0QI	A395BORC6	Karl	3	3	2	1.308E+09	Cough Medi	If you are looking for the secret ingredient		
6	5	B006K2ZZ7|	A1UQRSCLF	Michael D.		0	0	5	1.351E+09	Great taffy	Great taffy at a great price. There was a v		
7	6	B006K2ZZ7|	ADT0SRK1N	Twoapenny	0	0	4	1.342E+09	Nice Taffy	I got a wild hair for taffy and ordered this		
8	7	B006K2ZZ7|	A1SP2KVKF)	David C. Sul	0	0	5	1.34E+09	Great! Just	This saltwater taffy had great flavors and \		
9	8	B006K2ZZ7|	A3JRGQVEC	Pamela G. V	0	0	5	1.336E+09	Wonderful,	This taffy is so good. It is very soft and ch		
10	9	B000E7L2R4	A1MZYO9T7	R. James	1	1	5	1.322E+09	Yay Barley	Right now I'm mostly just sprouting this so		
11	10	B00171APV,	A21BT40VZ(	Carol A. Ret	0	0	5	1.351E+09	Healthy Dog	This is a very healthy dog food. Good for t		
12	11	B0001PB9FI	A3HDKO7O'	Canadian Fa	1	1	5	1.108E+09	The Best Ho	I don't know if it's the cactus or the tequila		
13	12	B0009XLVG(	A2725IB4YY	A Poeng "Sp	4	4	5	1.283E+09	My cats LO\	One of my boys needed to lose some weig		
14	13	B0009XLVG(	A327PGT22\	LT	1	1	1	1.345E+09	My Cats Are	My cats have been happily eating Felidae f		

II. NLP

After preprocessing, using NLP Techniques we expect to meet our goals.

Using resources such as VADER , we can do sentiment analysis and feature extraction.

III. User Outputs

Using some dashboards for users, we can present the summaries of the analysis and other details about the product reviews.

Challenges

- As beginners we are inexperienced and are still in the learning phase.
- Many reviews may have mixed reactions and may be unclear.
- It will be challenging to handle the dataset and ensure its quality.
- Some reviews may be multilingual.

Expected Impact

- Improved decision-making for customers through summarized and categorized reviews.
- Better business strategies for sellers by identifying product strengths and weaknesses.
- Enhanced trust in the review ecosystem by identifying and filtering fake reviews.