SYNOPSIS

OF

CAPSTONE PROJECT

ON

"SENTIMENT ANALYSIS OF CUSTOMER

REVIEW DATA"

PROJECT SUMMARY

Batch details	PGPDSE-FT Hyderabad July 2023
Team members	1. Bandi Nikhil 2. Kata Ganesh 3. Jupally Rishivedh 4. Palepu Vaishnavi 5. Pamarti Vibha 6. Reddy Manideep
Domain of Project	Marketing and Retail Analysis
Proposed project title	SENTIMENT ANALYSIS OF CUSTOMER REVIEW DATA
Group Number	4
Team Leader	-
Mentor Name	Mr. Manvendra Singh

Date:

Signature of the Mentor   Signature of the Team
Leader

Table of Contents

ABSTRACT

Sentiment analysis helps to determine the hidden intentions of the respective authors in each topic and provides an evaluation report on the polarity of each document. Polarity can be positive, negative or neutral. It can be seen that most of the data related to sentiment analysis includes feedback provided by different customers on any product. Therefore, the reviews can be correctly classified into any type of classification based on polarity to gain sufficient knowledge about the product. In this project, we propose an approach to classify dataset created based on sentiment analysis into different polar groups. The results obtained in relation to the accuracy values of the algorithm using different performance parameters applied on the dataset is critically examined.

The purpose of this project is to perform sentiment analysis of product-based reviews by using data.

This project uses online product reviews collected from "amazon.com". We expect to do Classification at the review level of survey data with promising results.

INTRODUCTION

Sentiment analysis, also known as opinion mining, analyses people's opinions and feelings about entities such as products, organizations, and their attributes. There are many social media on the web where you can easily get information about any product through reviews, blogs and comments. These comments not only help analysts study the reaction to any commodity or product, but may also give customers an opportunity to learn about demand in any given region. Various machine learning techniques are often used by researchers and professionals to perform analysis of people's emotions in an optimal manner.

Sentiment analysis examines how the text conveys an emotion. Customer feedback, survey responses, and product reviews are all frequently used. This is useful in a variety of situations, including social media monitoring, reputation management, and customer service. For example, analysing thousands of product reviews can provide important feedback on product pricing and features.

People's willingness to interact with a company and their overall perception of a brand are heavily influenced by public opinion. According to Podium research, 93% of shoppers believe that online reviews influence their purchasing decisions. After reading some negative reviews, users may not want to give the company a chance. They don't check if the feedback is genuine or not. They will choose another way. In this situation, companies that carefully monitor their reputation can quickly address issues and improve their operations based on feedback. In the age of information, such an analysis allows accurate measurement of people's attitudes towards companies.

There are two main categories of machine learning techniques that are mainly used in sentiment analysis: supervised learning and unsupervised learning. In supervised learning, datasets are labeled during analysis and trained to provide meaningful outputs that are useful for decision making. Unlike supervised learning, unsupervised learning does not provide labeled data, which is very difficult to process. Different clustering algorithms are often investigated to solve this problem.

This project highlights the effect of supervised learning techniques on labeled data.

A natural language processing (NLP) technique is used for determining whether data is positive, negative, or neutral. It is often used with contextual data to help organizations track brand and product sentiment in consumer feedback and better understand customer desires.

This tool helps businesses extract information from unstructured text found on the Internet, including emails, blog posts, support tickets, web chats, social media channels, forums, and comments.

Types of Sentiment Analysis

- Step-by-step sentiment analysis

If polarity accuracy is important to your business, expand your polarity category to include multiple positive and negative levels.

very positive

Positive

neutral

negative

very negative

This is commonly referred to as step-by-step or detailed sentiment analysis and can be used to interpret 5-star ratings in reviews. For example:

Very positive = 5 stars

Very negative = 1 star

- Recognizing emotions

Emotion Analysis Emotion Recognition allows you to recognize emotions such as happiness, frustration, anger and sadness across polarities.

Many emotion recognition systems use vocabularies (ie, lists of words and the emotions they convey) or complex machine learning algorithms.

One of the disadvantages of using a dictionary is that people express their feelings in different ways. Some words that usually express anger, such as bad and kill (e.g. your product is so bad

or your customer support is killing me) can also express happiness (e.g. this is terrible or you are killing it).

- <u>Aspect-based sentiment analysis</u>

Typically, when analyzing textual sentiment, you want to know which aspects or specific features people refer to positively, neutrally, or negatively.

This is where aspect-based sentiment analysis comes in handy. For example, in this product review, if you say, "This camera's battery life is too short," a facet-based classifier can determine that the sentence expresses a negative opinion about battery life. of the desired product

- <u>Multilingual sentiment analysis</u>

Multilingual sentiment analysis can be difficult. It requires a lot of preprocessing and resources. Most of these resources are available online (eg, emotion dictionaries), but others need to be developed (eg, translated corpora and noise detection algorithms). However, you need to know how to code to use them.

Alternatively, you can use the language classifier to automatically detect the language in text and train a custom sentiment analysis model to classify the text into the language of your choice.

ABOUT DOMAIN

Retail analytics focuses on providing insights related to sales, inventory, customers, and other important aspects crucial for merchants' decision-making process. The discipline encompasses several granular fields to create a broad picture of a retail business' health, and sales alongside overall areas for improvement and reinforcement.

Essentially, retail analytics is used to help make better choices, run businesses more efficiently, and deliver improved customer service analytics.

The field of retail analysis goes beyond superficial data analysis, using techniques like data mining and data discovery to sanitize datasets to produce actionable BI insights that can be applied in the short-term.

Moreover, companies use these analytics to create better snapshots of their target demographics. By harnessing sales data analysis, retailers can identify their ideal customers according to diverse categories such as age, preferences, buying patterns, location, and more.

Essentially, the field is focused not just on parsing data, but also defining what information is needed, how best to gather it, and most importantly, how it will be used.

Electronic commerce is buying and selling of goods or services through internet. The products are advertised on e-commerce websites developed by online retailers. These companies have domain expertise in the products and services they offer to buyers.

Various companies like Amazon, Big Bazaar, Flipkart, Snapdeal offer products that are essential to daily needs, from books to mobiles, electronic gadgets to clothing, groceries to kitchen essentials.

Various advantages eCommerce offers are:
- A comprehensive range of options for product with respect to their gender, age, brand, fashion, price range.
- Round-the-clock availability to select and purchase various items of desire and need.

- Payment options from net banking, credit cards, e-wallet, cash on delivery, etc.

- Privacy of personal information and guarantee of secured transactions.

- Refund option in case of return of goods if the user changes his/her mind from purchase.

- Once registered on portal Ease of login, view wish list, the status of the order placed, history of purchases, repeat order from the past purchases, option to save credit card details for quick pay, the intimation of availability of a product in case it is not in stock, in case product is in the wishlist.

- Options to purchase items from the web as well as mobile platforms for the e-commerce web portal.

Electronic commerce is buying and selling goods or services through the internet. The products are advertised on e-commerce websites developed by online retailers. Types of eCommerce portals are B2B, B2C, C2C. Some of the features like an exhaustive range of option for product to select from, 24 x7 availability of product to select desired items, options of paying like net banking, credit cards, cash on delivery, a guarantee of secured transactions, refund option in case of return of purchased items, wishlist, order status, quick pay options.

PROBLEM STATEMENT

"In the landscape of electronic product reviews, the challenge persists in effectively employing sentiment analysis to derive actionable insights. Despite the prevalence of sentiment analysis tools, accurately gauging the sentiments expressed in electronic product reviews remains a complex task. The issue revolves around the need for a more nuanced and contextually aware sentiment analysis approach that can decipher the subtleties and intricacies of user sentiments towards electronic products. Addressing this challenge involves developing or employing sentiment analysis models tailored to the specific jargon, technical aspects, and domain-specific language used in electronic product reviews. The objective is to create a more reliable and accurate sentiment analysis framework capable of capturing the diverse spectrum of opinions and emotions expressed in these reviews. Enhancing sentiment analysis in the realm of electronic product reviews is crucial for empowering consumers with trustworthy insights and supporting manufacturers in understanding user sentiments to improve product development and customer satisfaction."

CONTRIBUTION TO THE BUSINESS

- **Sorting Data at Scale**: Can you imagine manually sorting through thousands of tweets, customer support conversations, and surveys? There is not much business data to process manually. Sentiment analysis helps businesses process large amounts of unstructured data in an efficient and cost-effective manner.

- **Real-time analysis**: Sentiment analysis helps identify critical issues in real time. For example, is there an escalation of the public relations crisis in social media? Are angry customers about to defect?

- **Compatible standards**: It is estimated that people agree only about 60-65% of the time when judging the sentiment of a particular text. Labeling text with emotions is highly subjective and is influenced by personal experiences, thoughts and beliefs.

Review examples of sentiment analysis data

- **"Netflix has the best selection of films"**

- **"Hulu has a great UI"**

- **" I dislike like the new crime series"**

- **"I hate waiting for the next series to come out"**

PROBLEMS FACED

The most fundamental problem in sentiment analysis is the polarity of sentiments. Classification based on a dataset containing more than 5.1 million product reviews. Search Amazon.com for products in four categories.

The maximum entropy of the POS tag is used to classify words in a sentence. Additional Python programs to speed up the process. Negative words like no, that's not true - They are placed in adverbs, but they are negation of adjectives and negation of adjectives. Verbs are used specifically to identify expressions.

"The product didn't live upto my expectations …….. but it was good enough for the money paid"
The given line is a product review that states that "it" (the product) is not great but it was good enough at the particular price point. Understanding such sentiments require multiple tasks.

Hence, SENTIMENTAL ANALYSIS is a kind of text classification based on Sentimental Orientation (SO) of opinion they contain. Sentiment analysis of product reviews has recently become very popular in text mining and computational linguistics research.

• Firstly, evaluative terms expressing opinions must be extracted from the review.

• Secondly, the SO, or the polarity, of the opinions must be determined.

• Thirdly, the opinion strength, or the intensity, of an opinion should also be determined.

• Finally, the review is classified with respect to sentiment classes, such as Positive,Neutral and Negative, based on the SO of the opinions it contains

Below are the various classification models currently used in the industry:

1. Naïve Bayesian
2. Random Forest
3. Logistic Regression
4. Support Vector Machine.

Project Approach- MODELS USED FOR THIS PROJECT
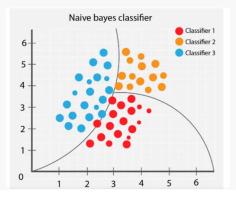
Naive Bayes sentiment analysis

It is called 'Naïve' because it uses the assumption that the occurrence of one feature is independent of other features. For instance, it identifies the orange fruit based on color, shape and taste with each feature independently being assessed to arrive at the conclusion. The 'Bayes' is because it is based on the principle of the Bayes theorem.

The Bayes theorem relies on the concept of conditional probability or **the probability that event A occurs when event B occurs.** The theorem in effect states that *t*he probability of A if B is true = the probability of B if A is true, multiplied by the times the probability of A being true and the whole divided by the probability of B being true*:*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In Naïve Bayes sentiment analysis, the Bayesian classifier classifies documents, text or products as positive or negative.
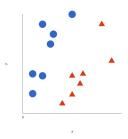
For example, in the sentence 'I like this product very much', you get a clear sense of the positive sentiment. The classifier calculates each probability value and the class is selected as **positive** because the positive value outweighs it.
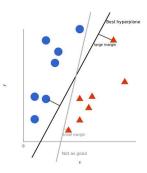
Support Vector Machines (SVM)

A support vector machine is another supervised machine learning model, similar to linear regression but more advanced. SVM uses algorithms to train and classify text within our sentiment polarity model, taking it a step beyond *X/Y* prediction.

For a simple visual explanation, we'll use two tags: *red* and *blue*, with two data features: *X* and *Y*. We'll train our classifier to output an *X/Y* coordinate as either *red* or *blue*.
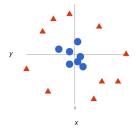

The SVM then assigns a hyperplane that best separates the tags. In two dimensions this is simply a line (like in linear regression). Anything on one side of the line is *red* and anything on the other side is *blue*. For sentiment analysis this would be *positive* and *negative*.

In order to maximize machine learning, the best hyperplane is the one with the largest distance between each tag:
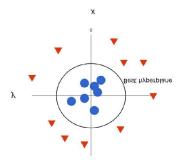

However, as data sets become more complex, it may not be possible to draw a single line to classify the data into two camps:

Using SVM, the more complex the data, the more accurate the predictor will become. Imagine the above in three dimensions, with a *Z* axis added, so it becomes a circle.
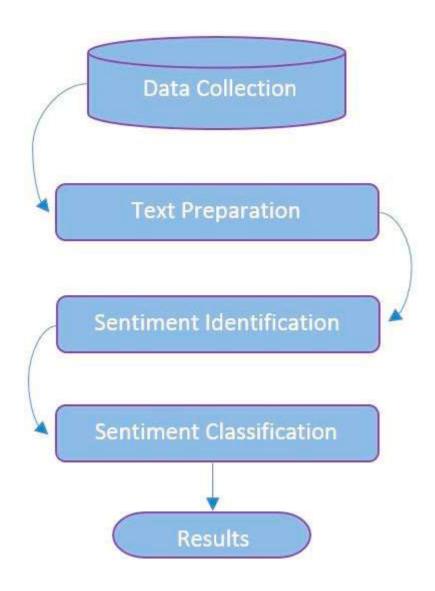
Mapped back to two dimensions with the best hyperplane, it looks like this:

Very simply put, SVM allows for more accurate machine learning because it's multidimensional.

OBJECTIVE OF THE PROJECT

- Scrapping product reviews featuring electronic products specifically amazon.com.
- Analyse and categorize review data.
- Analyse sentiment on dataset from document level (review level).
- Categorization or classification of opinion sentiment into-
    1. Positive
    2. Negative
    3. Neutral
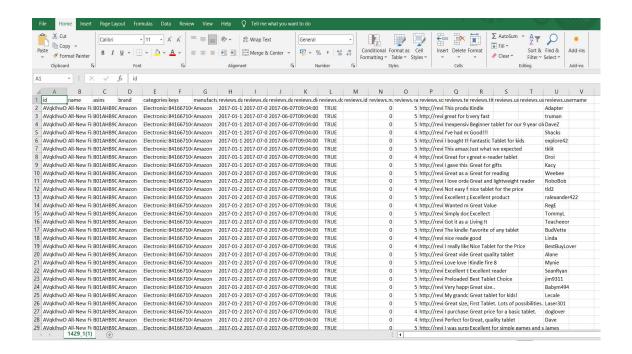
DATA USED

CUSTOMER REVIEWS OF AMAZON ELECTRONICS DATA

This is a list of over 34,000 consumer reviews for Amazon products like the Kindle, Fire TV Stick, and more provided by Datafiniti's Product Database. The dataset includes basic product information, rating, review text, and more for each product.

DATASET

https://drive.google.com/file/d/1nq_meLq4sqiPEnARj6zkMNXgY9DFSsIn/view?usp=drive_link

DATA INFORMATION

A description of each column based on the information provided:

1. id: An identifier, likely unique, representing an item or product.

2. name: The name or title of the product.

3. asins: ASINs (Amazon Standard Identification Numbers) associated with the product. ASINs are unique identifiers assigned by Amazon to products for sale.

4. brand: The brand or manufacturer of the product.

5. categories: Categories or product types the item belongs to.

6. keys: Keywords or specific attributes associated with the product.

7. manufacturer: The company or entity that produced or manufactured the item.

8. reviews.date: The date when the review was posted.

9. reviews.dateAdded: The date when the review was added to the dataset.

10. reviews.dateSeen: The date when the review was observed or seen.

11. reviews.didPurchase: Indicates whether the reviewer made a purchase of the product.

12. reviews.doRecommend: Indicates if the reviewer recommends the product.

13. reviews.id: ID associated with the review.

14. reviews.numHelpful: The number of helpful votes the review received.

15. reviews.rating: The rating given in the review (e.g., star rating).

16. reviews.sourceURLs: URLs or sources of the reviews.

17. reviews.text: The text content of the review.

18. reviews.title: The title or summary of the review.

19. reviews.userCity: The city of the reviewer (if provided).

20. reviews.userProvince: The province or state of the reviewer (if provided).

21. reviews.username: The username or identifier of the reviewer.

These columns seem to contain information about products, their reviews, and related metadata such as dates, user details, and review characteristics.

CONCLUSION

In this study, we proposed sentiment diffusion for customer review analysis. Suggested sentiment dissemination increased the consistency of existing sentiment analysis domains by reinforcing sentiment words in both context- and domain-specific ways. With the help of sentiment analysis implementation, we can conclude that most people are satisfied with the products and services they received from Amazon.

Sentiment analysis is a good process to gather insights into each product, which ultimately benefit present and future customers and e-commerce companies.

Sentiment analysis is important because, based on bad reviews, the e-commerce company makes those products better or replaces those products with better and newer ones, which ultimately improves the overall customer service too.

REFERENCES

What is Retail Analytics? | Sisense

Sentiment Analysis & Machine Learning (monkeylearn.com)

https://www.kaggle.com/code/amuhialdeen/lstm-in-amazon-customer-sentiments