European Economic Letters
ISSN 2323-5233
Vol 13, Issue 4 (2023)
https://doi.org/10.52783/eel.v13i4.691
http://eelet.org.uk

Exploratory Sentiment Analysis of Sales Data

1 Dr. Deepti Sharma, 2 Dr. Archana B.Saxena, 3 Dr. Deepshikha Aggarwal
Department of Information Technology
JIMS, Sec-5, Rohini, Delhi

Abstract

Sentiment analysis has numerous real-world applications, including Business and Market Intelligence, Customer Support, Social Media Monitoring, Political Analysis, Market Research. Companies use sentiment analysis to gauge public opinion about their products or services through customer reviews and social media data. Sentiment analysis can be used to automatically classify and prioritize customer support tickets based on the sentiment expressed by customers. This research deals with amazon sales data. It does exploratory data analysis, created word cloud for the data and also performs data validation.

Keywords: Sentiment analysis, exploratory data analysis, social media, model validation

1. Introduction

Exploratory sentiment analysis of sales data can provide valuable insights into customer satisfaction, product quality, and areas for improvement. By visualizing and understanding sentiment trends and patterns, businesses can make informed decisions to enhance their sales strategies and overall customer experience.

Sentiment analysis has numerous real-world applications, including Business and Market Intelligence, Customer Support, Social Media Monitoring, Political Analysis, Market Research. Companies use sentiment analysis to gauge public opinion about their products or services through customer reviews and social media data. Sentiment analysis can be used to automatically classify and prioritize customer support tickets based on the sentiment expressed by customers. It helps organizations track and analyse social media conversations to understand how their brand is perceived and respond to customer feedback. Sentiment analysis is used in politics to gauge public sentiment about politicians, policies, and issues. It aids in understanding consumer preferences and trends by analysing surveys and feedback.

Overall, sentiment analysis is a valuable tool for understanding and making sense of the vast amount of textual data available in the digital age, enabling organizations to make data-driven decisions and respond to customer needs more effectively. In this research, amazon sales data is used to analyse and do sentiment analysis for various products. Based on which, company can take further decisions for products.

2. Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique used to determine the sentiment or emotional tone expressed in a piece of text, such as a sentence, paragraph, or document. The primary goal of sentiment analysis is to automatically classify the text as positive, negative, or neutral based on the underlying sentiment conveyed by the words and phrases used.

Sentiment analysis typically takes text data as input. This can include social media posts, customer reviews, news articles, product descriptions, and more. The analysis involves classifying the text into one or more predefined sentiment categories. The most common categories are positive, negative, and neutral, but some applications may use more fine-grained categories like strongly positive, moderately positive, weakly positive. Before sentiment analysis can be performed, the text is often pre-processed to remove noise, such as punctuation, stop words, and special characters. Tokenization, stemming, and lemmatization are some of the common pre-processing techniques used. Sentiment analysis algorithms rely on various features extracted from the text, such as words, phrases, or even entire sentences. These features are used to train machine learning models or perform statistical analysis to determine sentiment. Different machine learning and NLP techniques can be used for sentiment analysis. These systems use predefined rules and dictionaries to identify sentiment-bearing words and phrases and calculate an overall sentiment score. Supervised learning models, such as support vector machines (SVM), logistic regression, and deep learning techniques like recurrent neural networks (RNNs) and transformers, are used to train models on labelled datasets to predict sentiment. The performance of a sentiment analysis system is assessed using various evaluation metrics, including accuracy, precision, recall, F1 score, and more, depending on the specific application. Sentiment analysis can play a significant role in analysing sales data and has several important implications for businesses. Sales data often includes customer reviews, feedback, and comments.

European Economic Letters
ISSN 2323-5233
Vol 13, Issue 4 (2023)
https://doi.org/10.52783/eel.v13i4.691
http://eelet.org.uk

Sentiment analysis can automatically process this textual data to determine whether the sentiments expressed by customers are positive, negative, or neutral. This feedback can be invaluable in understanding how customers perceive products or services. Sentiment analysis helps businesses identify specific pain points or concerns raised by customers. By analysing negative sentiment, companies can pinpoint areas that require improvement, whether it's product quality, customer service, or other aspects of the customer experience. Sentiment analysis can extend beyond your own sales data to include data from competitors. By analyzing customer sentiment toward your competitors, you can identify areas where your products or services excel and areas where they may fall short in comparison. This can inform your marketing and product development strategies. Sentiment analysis can provide insights into customer preferences and desires. By analyzing the sentiment associated with various product features or attributes, businesses can make informed decisions about which features to prioritize in future product development. By categorizing customers based on sentiment, businesses can create customer segments. For example, you can identify loyal customers who consistently express positive sentiment and target them with loyalty programs. Conversely, you can identify dissatisfied customers and take steps to retain them or address their concerns. Sentiment analysis can provide real-time insights into customer sentiment. This is particularly useful for businesses with e-commerce platforms, as they can monitor sentiment as it changes over time and respond quickly to emerging issues or trends. Sentiment analysis can be integrated into sales forecasting models. For example, if a product launch generates overwhelmingly positive sentiment, it may lead to higher sales than anticipated. Conversely, negative sentiment could signal potential sales challenges.

In summary, sentiment analysis is a valuable tool for businesses looking to gain deeper insights into their sales data. It enables companies to understand customer sentiment, identify areas for improvement, make data-driven decisions, and ultimately enhance their products, services, and marketing strategies to drive sales growth.

3. Methodology

a. Amazon sales data analysis

Analyzing Amazon sales data can provide valuable insights for businesses selling products on the platform. Following steps are performed on sales data analysis on Amazon.

1. Data Collection:
Gather your Amazon sales data, which may include information on product sales, customer reviews, ratings, pricing, and more. Amazon provides various reports and APIs for accessing this data.

2. Data Preparation:
Clean and preprocess the data to remove duplicates, handle missing values, and format it for analysis.

3. Exploratory Data Analysis (EDA):
Descriptive data analysis is done to understand the basic characteristics of your sales data, including sales volume, revenue, and customer metrics. Sales trends over time using line charts or time series plots to identify seasonal patterns or long-term trends are also visualized.

4. Customer Segmentation:
Segment your customer base based on various criteria such as purchase history, location, demographics, or behavior. Analyze the purchasing behavior of different customer segments to tailor marketing and product strategies accordingly.

5. Product Analysis:
Individual product performance, including sales volume, revenue, and customer reviews are also analysed. Product ratings and reviews are used to identify opportunities for product improvements or expansions.

6. Customer Reviews and Sentiment Analysis:
Sentiment analysis on customer reviews is analysed to gauge customer satisfaction. Company can use this analysis to identify common issues or complaints mentioned in negative reviews and take corrective actions.

7. Reporting and Visualization:
Reports and visualizations (e.g., dashboards) are shown to present the findings to key stakeholders in your organization.

Analyzing Amazon sales data is an ongoing process that can help businesses optimize their presence on the platform, improve customer experiences, and drive sales growth. Additionally, using business intelligence tools or data analysis software can streamline and automate many aspects of this analysis.

European Economic Letters
ISSN 2323-5233
Vol 13, Issue 4 (2023)
https://doi.org/10.52783/eel.v13i4.691
http://eelet.org.uk

a. Pre-processing of data

Data preprocessing is a crucial step in data analysis and machine learning. It involves the cleaning and transformation of raw data into a format that is suitable for analysis or modeling. Proper data preprocessing helps improve the quality of the data and ensures that the data is ready for the specific tasks you want to perform, such as machine learning, statistical analysis, or data visualization.

b. Feature extraction

Feature extraction is a process in data analysis and machine learning where relevant information or features are selected or transformed from the raw data to create a more compact and meaningful representation for further analysis or modeling. This step is crucial because it helps reduce the dimensionality of the data, improve model performance, and capture the most important patterns or characteristics.

Feature extraction is a critical step in the data preprocessing pipeline and can significantly affect the quality of the results obtained from data analysis or machine learning models. The choice of feature extraction techniques should be guided by the nature of the data, the problem domain, and the specific goals of the analysis or modeling task.

c. Sentiment Analysis with Python

Sentiment analysis can be performed in Python using various libraries and tools.

Install Required Libraries: First, make sure you have the necessary Python libraries installed. If not there, they can be installed using pip command.

Import Libraries: In your Python script or Jupyter Notebook, import the required libraries:

Download NLTK Data (if not already downloaded): If NLTK data is not already downloaded, you can do so with the following commands.

Load or Collect Text Data: Load the data that you want to analyze for sentiment. This can be tweets, reviews, comments, or any other text data source.

Perform Sentiment Analysis:

Using TextBlob (a simplified approach)**:**
TextBlob is a simple library that provides a straightforward way to perform sentiment analysis. The sentiment score or compound score indicates the sentiment of the text. A positive score suggests positive sentiment, a negative score suggests negative sentiment, and scores close to zero suggest neutral sentiment.

Following steps are involved during analysis of amazon data.

a. Import relevant libraries.

```
import nltk
nltk.download('punkt')
nltk.download('stopwords')
from nltk.corpus import stopwords
```

European Economic Letters
ISSN 2323-5233
Vol 13, Issue 4 (2023)
https://doi.org/10.52783/eel.v13i4.691
http://eelet.org.uk

b. Load data

```
data = pd.read_csv('amazonreviewonly.csv')
data.head()
```

	sentiment	review_title
0	1	Satisfied,Charging is really fast,Value for mo...
1	1	A Good Braided Cable for Your Type C Device,Go...
2	0	Good speed for earlier versions,Good Product,W...
3	1	Good product,Good one,Nice,Really nice product...
4	1	As good as original,Decent,Good one for second...

c. Pre-processing of data

```
data.loc[data['sentiment']<=4,'sentiment'] = 0
data.loc[data['sentiment']>4,'sentiment'] = 1
```

```
stp_words=stopwords.words('english')
def clean_review(review):
  cleanreview=" ".join(word for word in review.
                       split() if word not in stp_words)
  return cleanreview

data['review_title']=data['review_title'].apply(clean_review)
```

d. Generating Word Cloud

```
consolidated=' '.join(word for word in data['review_title'][data['sentiment']==0

wordCloud=WordCloud(width=1600,height=800,random_state=21,max_font_size=110)
plt.figure(figsize=(15,10))
plt.imshow(wordCloud.generate(consolidated),interpolation='bilinear')
plt.axis('off')
plt.show()
```

European Economic Letters
ISSN 2323-5233
Vol 13, Issue 4 (2023)
https://doi.org/10.52783/eel.v13i4.691
http://eelet.org.uk

e. Training and Testing a model

```
cv = TfidfVectorizer(max_features=2500)
X = cv.fit_transform(data['review_title'] ).toarray()
```

```
from sklearn.model_selection import train_test_split
x_train ,x_test,y_train,y_test=train_test_split(X,data['sentiment'],
                                                test_size=0.25 ,
                                                random_state=42)
```

f. Experimentation for Model Validation

```
from sklearn.linear_model import LogisticRegression

model=LogisticRegression()

#Model fitting
model.fit(x_train,y_train)

#testing the model
pred=model.predict(x_test)

#model accuracy
print(accuracy_score(y_test,pred))
```

Output :

```
0.81632
```

The above output shows the accuracy as 81%. The model is trained and tested on the data.

References:

[1] A.Pak and P. Paroubek. „Twitter as a Corpus for Sentiment Analysis and Opinion Mining". In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2010, pp.1320-1326

[2] R. Parikh and M. Movassate, “Sentiment Analysis of User- GeneratedTwitter Updates using Various Classi_cation Techniques",CS224N Final Report, 2009

[3] Pablo Gamallo, Marcos Garcia, “Citius: A Naive-Bayes Strategyfor Sentiment Analysis on English Tweets", 8th InternationalWorkshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland,Aug 23-24 2014, pp 171-175

[4]] Neethu M,S and Rajashree R,” Sentiment Analysis using Machine Learning Techniques” 4th ICCCNT 2013,at Tiruchengode, India. IEEE – 31661

[5] D Aggarwal, D Sharma , Application of clustering for student result analysis, - Int J Recent Technology and Enggineering, 2019.

[6] D Sharma, D Aggarwal, A Gupta , A study of consumer perception towards mwallets - Int. J. Sci. Technology, 2019

[7]] V. M. K. Peddinti and P. Chintalapoodi, “Domain adaptation in sentiment analysis of twitter,” in Analyzing Microtext Workshop, AAAI, 2011.