

SENTIMENT ANALYSIS FOR MARKETING

AI PHASE 5

PROBLEM DEFINITION:

The problem at hand is to perform sentiment analysis on customer feedback in order to gain valuable insights into competitor products. By understanding customer sentiments, companies can identify strengths and weaknesses in competing products, which can help improve their own offerings. This project requires the utilization of various Natural Language Processing (NLP) methods to extract meaningful insights from customer feedback.

DESIGN THINKING PROCESS:

1. Empathize:

- ✓ Understand the dataset and its attributes
- ✓ Recognize the importance of understanding customer sentiments for improving products and gaining a competitive edge.

2. Define:

- ✓ Define the problem: performing sentiment analysis on customer feedback for competitor products.
- ✓ Set specific objectives, such as identifying sentiment distribution and using NLP methods for insights.

3. Ideate:

- ✓ Use NLP techniques and tools suitable for sentiment analysis, such as VADER, NLTK, or Transformers.
- ✓ Consider data sources, including customer reviews, social media, and surveys.
- ✓ Explore potential features to analyze, like sentiment labels, common keywords, and negative feedback reasons.

4. Prototype:

- ✓ Create a prototype or proof-of-concept code to perform sentiment analysis on customer feedback data.
- ✓ Test and refine NLP methods and parameters to ensure accurate sentiment categorization.
- ✓ Visualize the sentiment distribution and common themes.

5. Test:

- ✓ Apply the prototype to a larger dataset of competitor product feedback.
- ✓ Evaluate the accuracy of sentiment categorization and the quality of insights obtained.

6. Implement:

- ✓ Develop a full-scale sentiment analysis solution using the selected NLP methods.
- ✓ Integrate the solution with data sources to regularly analyze customer feedback.

PHASES OF DEVELOPMENT:

Data Preprocessing:

- ✓ Load the dataset and remove unnecessary columns.
- ✓ Handle the missing values in the dataset which can be problematic for data analysis.
- ✓ Convert the sentiment labels and text to lowercase to ensure uniformity.
- ✓ Verify the data changes for the further analysis.

Sentiment Analysis Techniques:

- ✓ Used VADER sentiment analysis tool to perform sentiment analysis on the preprocessed text and generated a graph for sentiment distribution.
- ✓ Report summary is generated using RoBERTa model

DATASET DESCRIPTION:

This data contains information about the customer feedback and sentiment analysis of a airline with attributes like tweet id, airline sentiment, negative reasons. It comprises of 14640 rows and 15 attributes.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	tweet_id	airline_ser	airline_ser	negativere	negativere	airline	airline_ser	name	negativere	retweet_c	text	tweet_coc	tweet_cre	tweet_loc	user_timezone		
2	5.70E+17	neutral	1			Virgin America	cairdin			0	@VirginAmerica Wha #####				Eastern Time (US & Canada)		
3	5.70E+17	positive	0.3486			0 Virgin America	jnardino			0	@VirginAmerica plus ' #####				Pacific Time (US & Canada)		
4	5.70E+17	neutral	0.6837			Virgin America	yvonnalynn			0	@VirginAmerica I didi #####	Lets Play			Central Time (US & Canada)		
5	5.70E+17	negative	1	Bad Flight	0.7033	Virgin America	jnardino			0	@VirginAmerica it's re #####				Pacific Time (US & Canada)		
6	5.70E+17	negative	1	Can't Tell	1	Virgin America	jnardino			0	@VirginAmerica and i #####				Pacific Time (US & Canada)		
7	5.70E+17	negative	1	Can't Tell	0.6842	Virgin America	jnardino			0	@VirginA #####				Pacific Time (US & Canada)		
8	5.70E+17	positive	0.6745			0 Virgin America	cjmginis			0	@VirginAmerica yes, t #####	San Franci			Pacific Time (US & Canada)		
9	5.70E+17	neutral	0.634			Virgin America	pilot			0	@VirginAmerica Reall #####	Los Angele			Pacific Time (US & Canada)		
10	5.70E+17	positive	0.6559			Virgin America	dhepburn			0	@virginamerica Well, #####	San Diego			Pacific Time (US & Canada)		
11	5.70E+17	positive	1			Virgin America	YupitsTate			0	@VirginAmerica it wa #####	Los Angele			Eastern Time (US & Canada)		
12	5.70E+17	neutral	0.6769			0 Virgin America	idk_but_youtube			0	@VirginAmerica did y #####	1/1 loner s			Eastern Time (US & Canada)		
13	5.70E+17	positive	1			Virgin America	HyperCamiLax			0	@VirginAmerica I < #####	NYC			America/New_York		
14	5.70E+17	positive	1			Virgin America	HyperCamiLax			0	@VirginAmerica This i #####	NYC			America/New_York		
15	5.70E+17	positive	0.6451			Virgin America	mollanderson			0	@VirginAmerica @vir #####				Eastern Time (US & Canada)		
16	5.70E+17	positive	1			Virgin America	sjespers			0	@VirginAmerica Than #####	San Franci			Pacific Time (US & Canada)		

DATA PREPROCESSING STEPS:

Remove Unnecessary Columns

- ✓ Identify and remove columns from the dataset that are not needed
- ✓ for analysis. These columns might contain information that is irrelevant or redundant
- ✓ for specific task.
- ✓ By doing this, the dataset becomes more concise and focused.

Handle Missing Values

- ✓ check for missing values in specific columns. If there are any rows with missing values in the columns, remove those rows from the dataset.

Convert Sentiment Labels and text to Lowercase:

- ✓ To ensure consistency and to avoid issues related to letter casing, convert the text to lowercase.
- ✓ This ensures uniformity in sentiment labels.

Verify Data Changes:

- ✓ verify that the changes have been correctly applied to the dataset and ready for further analysis.

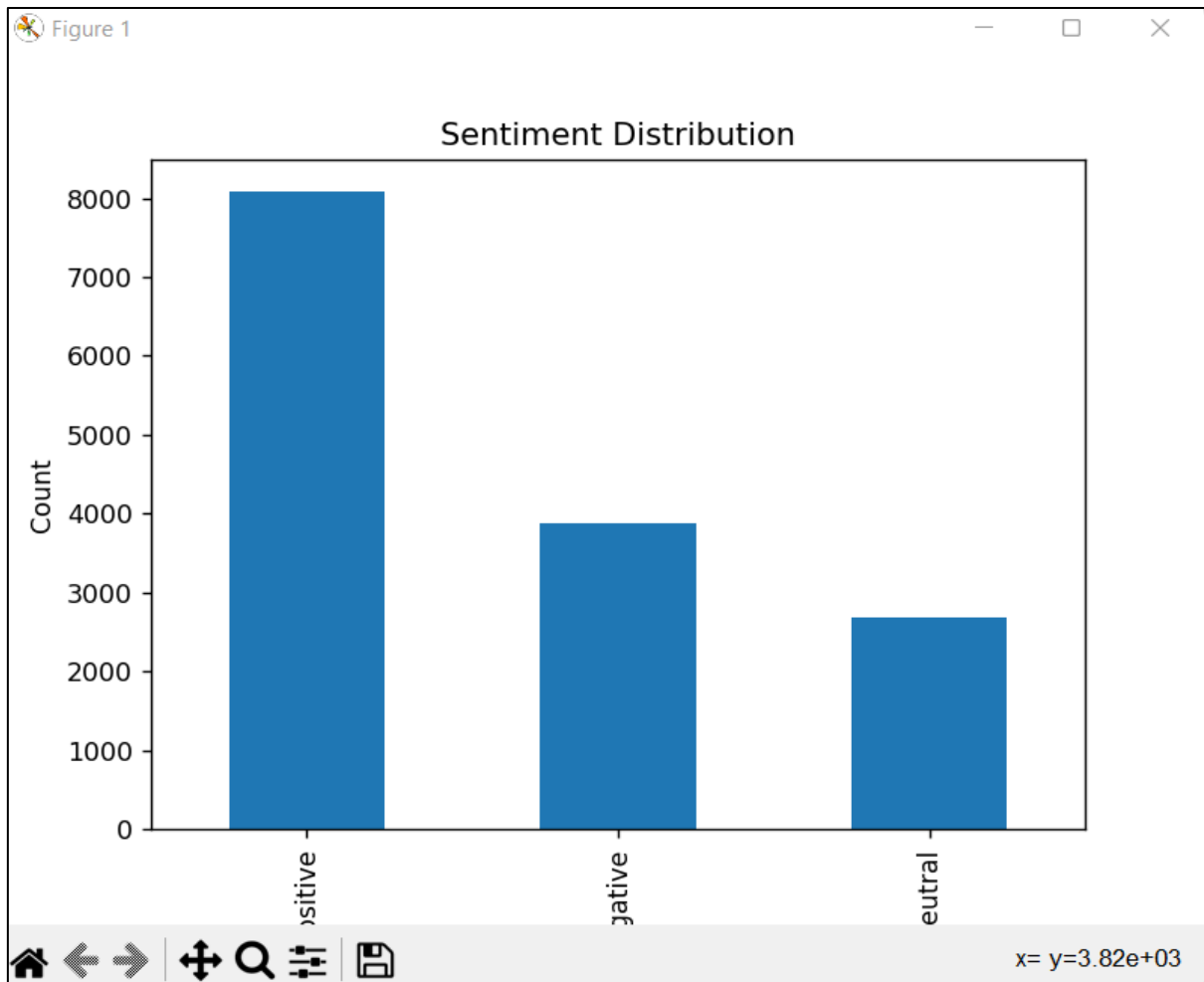
	airline_sentiment	airline_sentiment_confidence	negativereason	negativereason_confidence	text
0	neutral	1.0000	NaN	NaN	@virginamerica what @dhepburn said.
1	positive	0.3486	NaN	0.0000	@virginamerica plus you've added commercials t...
2	neutral	0.6837	NaN	NaN	@virginamerica i didn't today... must mean i n...
3	negative	1.0000	Bad Flight	0.7033	@virginamerica it's really aggressive to blast...
4	negative	1.0000	Can't Tell	1.0000	@virginamerica and it's a really big bad thing...
5	negative	1.0000	Can't Tell	0.6842	@virginamerica seriously would pay \$30 a fligh...
6	positive	0.6745	NaN	0.0000	@virginamerica yes, nearly every time i fly vx...
7	neutral	0.6340	NaN	NaN	@virginamerica really missed a prime opportuni...
8	positive	0.6559	NaN	NaN	@virginamerica well, i didn't..but now i do! :-d
9	positive	1.0000	NaN	NaN	@virginamerica it was amazing, and arrived an ...

PS C:\Users\RAMYA S\Desktop\My project>

SENTIMENT ANALYSIS TECHNIQUES:

VADER Sentiment Analysis Tool:

- ✓ The VADER sentiment analysis tool from NLTK (Natural Language Toolkit) to perform sentiment analysis on the preprocessed text of the tweets in the dataset.
- ✓ The sentiment labels by VADER tool are used for subsequent analysis and visualization to understand the distribution of sentiments in the dataset.



RoBERTa Model:

- ✓ RoBERTa model is used to generate text for the feedback report.
- ✓ The Hugging Face transformers library is used to initialize a text generation pipeline with RoBERTa as the underlying model.
- ✓ The "roberta-base" model is a pre-trained RoBERTa model that can generate coherent and contextually relevant text based on provided prompts.
- ✓ The generated text using Roberta model includes the summary report, basic statistics, and the summary of negative feedback reasons.

File	Edit	View
Summary Report:		
Total Rows: 14640		
Positive Feedback Rows: 2363		
Negative Feedback Rows: 9178		
Neutral Feedback Rows: 3099		
Reason and Summaries :		
+-----+-----+		
Reason	Summary	
+=====+		
Customer Service Issue	2910 passengers mentioned this reason.	
+-----+-----+		
Late Flight	1665 passengers mentioned this reason.	
+-----+-----+		
Can't Tell	1190 passengers mentioned this reason.	
+-----+-----+		
Cancelled Flight	847 passengers mentioned this reason.	
+-----+-----+		
Lost Luggage	724 passengers mentioned this reason.	
+-----+-----+		
Bad Flight	580 passengers mentioned this reason.	
+-----+-----+		
Flight Booking Problems	529 passengers mentioned this reason.	
+-----+-----+		
Flight Attendant Complaints	481 passengers mentioned this reason.	
+-----+-----+		
longlines	178 passengers mentioned this reason.	
+-----+-----+		
Damaged Luggage	74 passengers mentioned this reason.	
+-----+-----+		

CONCLUSION:

Sentiment analysis equips businesses with actionable insights to enhance customer satisfaction and address issues effectively. The sentiment analysis revealed a diverse range of sentiments in the dataset, with a notable number of negative comments. By categorizing feedback as positive, negative, or neutral and identifying specific negative reasons, companies can understand customer sentiment, recognize strengths and weaknesses in their offerings, and make data-driven improvements to enhance customer satisfaction and competitiveness.