

Project 5: Sentiment analysis for marketing

Project Title: *Sentiment Analysis*

PHASE 1 : Problem definition and design thinking

Problem statement:

This type of project can show you what it's like to work as an NLP specialist. For this project, you want to find out how customers evaluate competitor products, i.e., what they like and dislike. It's a great business case. Learning what customers like about competing products can be a great way to improve your own product, so this is something that many companies are actively trying to do. Employ different NLP methods to get a deeper understanding of customer feedback and opinion.

Problem definition:

The problem is to perform sentiment analysis on customer feedback to gain insights into competitor products. By understanding customer sentiments, companies can identify strengths and weaknesses in competing products, thereby improving their own offerings. This project requires utilizing various NLP methods to extract valuable insights from customer feedback.

Submission by : SURYA P ,

Department of Electronics & Communication, Anna university RC Coimbatore.

← → ↺

kaggle.com/zeus24/ibm-nm-ai-sentianalysis/edit

🔍 📄 ☆ 🌟 ⌵ 🏠 ⚙️

☰

IBM_NM_AI_sentianalysis Draft saved

File Edit View Run Add-ons Help

+ 🗑️ ✂️ 📄 📋 ▶ ▶▶ Run All Code ▾

● Draft Session (47m) H D D C P U R A M ⋮

+ 🔍 🏆 📊 🌐 <> 💬 🎓 ▾

```
for filename in filenames:
    print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved as ou
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the cu
```

[1]:

```
import pandas as pd
```

The pandas package is imported using the variable pd

```
[2]: data = pd.read_csv('/kaggle/input/twitter-airline-sentiment/Tweets.csv')
```

```
print(data)
```

	tweet_id	airline_sentiment	airline_sentiment_confidence	\
0	570306133677760513	neutral	1.0000	

Share Save Version 0

Notebook

Data

+ Add Data

Input

twitter-airline-sentiment

- Tweets.csv

database.sqlite


- Tweets

Output (56KB / 19.5GB)

/kaggle/working

Models

+ Add Models



2

Problem statement explanation:

- The project's objective is to perform sentiment analysis on customer feedback regarding competitor products.
- By utilizing various NLP methods, the aim is to extract valuable insights into customer sentiments, allowing companies to identify strengths and weaknesses in competing products.
- These insights enable data-driven decision-making, influencing product development, marketing strategies, and ultimately, improving overall customer satisfaction and competitive advantage.
- Continuous monitoring of evolving sentiments is essential for staying responsive to changing customer perceptions.

WHAT IS SENTIMENTAL ANALYSIS?

- Sentiment analysis is a Natural Language Processing (NLP) task that involves determining the sentiment or emotional tone expressed in a piece of text (Involves understanding emotions through symbolic expressions and text data)

PROJECT OBJECTIVE: To perform sentimental analysis on the given dataset – **twitter US airlines sentiment** to guide the company's decision

COLLECTION OF DATASET AND PRE-PROCESSING:

- The data is collected from the allotted dataset -- Twitter US Airline Sentiment { ***SOURCE : KAGGLE*** }
- We analyse the dataset for unfilled/empty responses.
- Either the given data can be structured into a proper form or we can deploy ML and deep learning to process the unstructured data.
- The dataset is imputed for the empty spaces using statistical techniques like –
Mean - Primarily for numerical data
Median and Mode – for Textual and other forms of data

AutoSaveOff

Tweets.csv • Saved to this PC

Search

Surya P

FileHomeInsertPage LayoutFormulasDataReviewViewHelp

CutCopyFormat Painter

Clipboard

Calibri11

B I U

Font

Wrap TextMerge & Center

Alignment

General

Number

Conditional FormattingFormat as Cell Styles

Styles

InsertDeleteFormat

Cells

AutoSumFillClear

Editing

Sort & Find & FilterSelect

Editing

CommentsShare

Add-ins

I389

California

	A	B	C	D	E	F	G	H	I	J	K
	tweet_id	airline_sentiment	_sentiment_confid	negative_reason	negative_reason_confidence	airline	name	text	tweet_location	user_timezone	
1	5.70306E+17	neutral	1			Virgin America	cardin	@VirginAmerica What @dhepburn said.			
2	5.70301E+17	positive	0.3486		0	Virgin America	jardino	@VirginAmerica plus you've added commercials to the experience... tacky.		Eastern Time (US & Canada)	
3	5.70301E+17	neutral	0.6837			Virgin America	yvonnyalynn	@VirginAmerica I didn't today... Must mean I need to take another trip!		Pacific Time (US & Canada)	
4	5.70301E+17	negative	1	Bad Flight	0.7033	Virgin America	jardino	@VirginAmerica it's really aggressive to blast obnoxious "entertainment" in your guests' faces & they have little recourse	Lets Play	Central Time (US & Canada)	
5	5.70301E+17	negative	1	Can't Tell	1	Virgin America	jardino	@VirginAmerica and it's a really big bad thing about it		Pacific Time (US & Canada)	
6	5.70301E+17	negative	1	Can't Tell	0.6842	Virgin America	jardino	@VirginAmerica seriously would pay \$30 a flight for seats that didn't have this playing.		Pacific Time (US & Canada)	
7								it's really the only bad thing about flying VA			
8	5.70301E+17	positive	0.6745		0	Virgin America	cjmoginnis	@VirginAmerica yes, nearly every time I fly VX this @keear woma! won't go away :)	San Francisco CA	Pacific Time (US & Canada)	
9	5.703E+17	neutral	0.634			Virgin America	thepburn	@VirginAmerica Really missed a prime opportunity for Men Without Hats parody, there. https://t.co/mVpG7grEzP	Los Angeles	Pacific Time (US & Canada)	
10	5.703E+17	positive	0.6553			Virgin America	Yupists	@VirginAmerica I well, I didn't about NDV I DO! :D	San Diego	Pacific Time (US & Canada)	
11	5.70235E+17	positive	1			Virgin America	idk_but_youtube	@VirginAmerica it was amazing, and arrived an hour early. You're too good to me.	Los Angeles	Eastern Time (US & Canada)	
12	5.70234E+17	positive	0.6763		0	Virgin America	HyperCamLax	@VirginAmerica did you know that suicide is the second leading cause of death among teens 10-24	Y11 lower squad	Eastern Time (US & Canada)	
13	5.7023E+17	positive	1			Virgin America	sjespers	@VirginAmerica I < 3 pretty graphics... so much better than minimal iconography :D	NYC	America/New_York	
14	5.7023E+17	positive	1			Virgin America	HyperCamLax	@VirginAmerica This is such a great deal! Already thinking about my 2nd trip to @Australia & I haven't even gone on my 1st trip yet! :p	NYC	America/New_York	
15	5.70287E+17	positive	0.6451			Virgin America	mollanderson	@VirginAmerica @virginmedia I'm flying your #fabulous #Seductive skies again! U take all the #stress away from travel http://t.co/ah0XtHKKjyn		Eastern Time (US & Canada)	
16	5.70286E+17	positive	1			Virgin America	smatruatermelon	@VirginAmerica @virginmedia I'm flying your #fabulous #Seductive skies again! U take all the #stress away from travel http://t.co/ah0XtHKKjyn	San Francisco, CA	Pacific Time (US & Canada)	
17	5.70282E+17	negative	0.6842	Late Flight	0.3684	Virgin America	ItzBianHunty	@VirginAmerica SFO-PDX schedule is still MIA	palo alto, ca	Pacific Time (US & Canada)	
18	5.7027E+17	positive	1			Virgin America	heatherovieda	@VirginAmerica So excited for my first cross country flight LAX to MCO I've heard nothing but great things about Virgin America. #23DaysToGo	west covina	Pacific Time (US & Canada)	
19	5.70277E+17	negative	1	Bad Flight	1	Virgin America	thebrandray	@VirginAmerica I flew from NYC to SFO last week and couldn't fully sit in my seat due to two large gentleman on either side of me. HELP!	this place called NYC	Eastern Time (US & Canada)	
20	5.7027E+17	positive	1			Virgin America	JMLpierce	I < 3, flying @VirginAmerica. < 3 < 3	Somewhere celebrating life.	Atlantic Time (Canada)	
21	5.70266E+17	positive	1			Virgin America	MISSGJ	@VirginAmerica you know what would be amazingly awesome? BOS-FLL PLEASE!!!!!! I want to fly with only you.	Quito		
22	5.70266E+17	negative	0.6705	Can't Tell	0.3614	Virgin America	DT_Les	@VirginAmerica why are your first fares in May over three times more than other carriers when all seats are available to select???	Boston Walham		
23	5.70264E+17	positive	1			Virgin America	ElvinaBeck	@VirginAmerica I love this graphic. http://t.co/UT5SGRvAaA			
24	5.70259E+17	positive	1			Virgin America	rjynoch21086	@VirginAmerica I love the hipster innovation. You are a feel good brand.	Los Angeles	Pacific Time (US & Canada)	
25	5.70259E+17	neutral	1			Virgin America	sweevicklee	@VirginAmerica will you be making BOS>LAS non stop permanently anytime soon?	Boston, MA	Eastern Time (US & Canada)	
26	5.70257E+17	negative	1	Customer Service Issue	0.3557	Virgin America	Leora3	@VirginAmerica you guys messed up my seating. I reserved seating with my friends and you guys gave my seat away... < 3 I want free internet	TM	Mountain Time (US & Canada)	
27	5.70249E+17	negative	1	Customer Service Issue	1	Virgin America	meredithjynn	@VirginAmerica status match program. I applied and it's been three weeks. Called and emailed with no response.			
28	5.7024E+17	negative	1	Can't Tell	0.6614	Virgin America	AdanSinger	@VirginAmerica what happened 2 ur vegan food options? At least say on site so I know I won't be able 2 eat anything for next 6 hrs #fail			
29	5.7021E+17	neutral	0.6854			Virgin America	blackjackpro311	@VirginAmerica do you miss me? Don't worry you'll be there very soon.	San Francisco, CA	Central Time (US & Canada)	
30	5.7020E+17	negative	1	Bad Flight	1	Virgin America	TenantsUpstairs	@VirginAmerica amazing to me that we can't get any cold air from the vents. #VX358 #noair #worstflightever #roasted #SFOtoBOS	San Mateo, CA & Las Vegas, NV		
31	5.70125E+17	neutral	0.615		0	Virgin America	Jervanttezz	@VirginAmerica LAX to EWR - Middle seat on a red eye. Such a noob maneuver. #zendambien #andchemix	Brooklyn	Atlantic Time (Canada)	
32	5.70114E+17	negative	1	Flight Booking Problems	1	Virgin America	Cuschool1	@VirginAmerica h! I just bled a cool birthday trip with you, but i can't add my elevate no. cause i entered my middle name during Flight Booking Problems < 3	Vienna		
33	5.70095E+17	neutral	1			Virgin America	amanduhmccarty	@VirginAmerica Are the hours of operation for the Club at SFO that are posted online current?	California, San Francisco	Pacific Time (US & Canada)	
34	5.70088E+17	negative	1	Customer Service Issue	1	Virgin America	NorthIsHomeTeam	@VirginAmerica help, left expensive headphones on flight 89 IAD to LAX today. Seat 2A. No one answering I&gt;number at LAX!	Washington DC	Quito	
35	5.70085E+17	negative	1	Customer Service Issue	1	Virgin America	miaerolinea	@VirginAmerica availing my return phone call, just would prefer to use your online self-service option :)		Pacific Time (US & Canada)	
36	5.70077E+17	positive	1			Virgin America	Nicplace	@VirginAmerica this is great news! America could start flights to Hawaii by end of year http://t.co/8p22j3te4 via @Pacifiobiznews	Texas	Central Time (US & Canada)	
37	5.70052E+17	neutral	0.6207			Virgin America	Nicplace	Nice RT @VirginAmerica: Vibe with the moodlight from takeoff to touchdown. #MoodlitMonday #ScienceBehindTheExperience http://t.co/Y700uNkTOP	Worldwide	Caracas	
38	5.70051E+17	positive	1			Virgin America	Nicplace	@VirginAmerica Moodlighting is the only way to fly! Best experience EVER! Cool and calming. < 3 < 3 < 3 #MoodlitMonday	Central Texas		
39	5.70045E+17	positive	1			Virgin America	Nicplace	@VirginAmerica @freddieavards Done and done! Best airline around, hands down!	Central Texas		
40	5.70039E+17	neutral	0.6791		0	Virgin America	elisha_malulani	@VirginAmerica when can I book my flight to Hav ai?	i'm creating a monster	Pacific Time (US & Canada)	
41	5.70038E+17	negative	1	Customer Service Issue	1	Virgin America	Jamesferandini	@VirginAmerica Your chat support is not working on your site. http://t.co/whp2GdVPLk	San Francisco, CA	Pacific Time (US & Canada)	
42	5.70034E+17	positive	0.6639			Virgin America	villencenzi	@VirginAmerica View of downtown Los Angeles, the Hollywood Sign, and beyond that rain in the mountains! http://t.co/Dv5r0lbtv			
43	5.70025E+17	negative	0.6688	Flight Booking Problems	0.6688	Virgin America	GottAmanda	@VirginAmerica Hey, first time flyer next week - excited! But I'm having a hard time getting my flights added to my Elevate account. Help?	Iowa City	Central Time (US & Canada)	
44	5.70016E+17	neutral	1			Virgin America	KGervaise	@VirginAmerica plz help me win my bid upgrade for my flight 2127 LAX-->SEA!! < 3 < 3 < 3	Los Angeles		
45	5.70015E+17	neutral	0.6578		0	Virgin America	papamurat	@VirginAmerica I have an unused ticket but moved to a new city where you don't fly. How can I fly with you before it expires? #travelhelp	Georgia	Pacific Time (US & Canada)	
46	5.70014E+17	neutral	1			Virgin America	arieldale	@VirginAmerica I have flights leaving Dallas for Seattle on time Feb 24?			
47	5.70012E+17	positive	1			Virgin America	vacations7	@VirginAmerica I'm @elevategold for a good reason: you rock!!	Los Angeles		
48	5.70011E+17	neutral	0.6793			Virgin America	Chelseapoe666	@VirginAmerica DREAM http://t.co/colA2dRfAoQ2 http://t.co/WdAc2Kfxk	Turks and caicos		
49	5.70011E+17	positive	1			Virgin America	Isaalko	@VirginAmerica I'm @elevategold for a good reason: you rock!!	Oakland via Midwest	Atlantic Time (Canada)	
50	5.70011E+17	neutral	1			Virgin America	granbrownie	@VirginAmerica @ladygaga @carriewood After last night #tribute #SoundOfMusic #Oscars2015 @ladygaga! I think @carriewood agree	New York, NY	Eastern Time (US & Canada)	
51	5.7001E+17	neutral	0.6436			Virgin America	granbrownie	@VirginAmerica @ladygaga @carriewood All were entertaining			
52	5.70003E+17	neutral	0.6764		0	Virgin America	granbrownie	@VirginAmerica Is flight 763 on it's way? Was supposed to take off 30 minutes ago. Website still shows "On Time" not "In Flight". Thanks.	Worldwide	Central Time (US & Canada)	
53	5.70007E+17	positive	0.657			Virgin America	granbrownie	@VirginAmerica @ladygaga @carriewood Julie Andrews all the way though @ladygaga was very impressive! ND to @Carriewood	Northern Virginia	Eastern Time (US & Canada)	
54	5.70004E+17	neutral	1			Virgin America	2v	@VirginAmerica with you flew out of Atlanta... Soon?	Los Angeles / Atlanta	Eastern Time (US & Canada)	

Tweets

55%

IBM_NM_AI_sentianalysis Draft saved

File Edit View Run Add-ons Help

Run All Code

Draft Session (1h:17m)

We now identify the columns that have blank responses.

```
data["airline"].drop_duplicates()
```

```
[17]: 0      Virgin America
      594      United
      4326     Southwest
      6746      Delta
      8966     US Airways
      11879    American
      Name: airline, dtype: object
```

VIRGIN AMERICA, UNITED, SOUTHWEST, DELTA, US AIRWAYS, AMERICAN are the airlines.

```
data["airline_sentiment_confidence"].drop_duplicates()
```

```
[18]: 0      1.0000
      1      0.3486
      2      0.6837
      6      0.6745
      7      0.6340
      ...
      14562   0.7257
      14569   0.7241
      14587   0.6384
      14594   0.7094
      14635   0.3487
      Name: airline_sentiment_confidence, Length: 1023, dtype: float64
```

```
data["airline_sentiment"].drop_duplicates()
```

```
[19]: 0      neutral
      1      positive
      3      negative
      Name: airline_sentiment, dtype: object
```

Notebook

Data

+ Add Data

Input

- twitter-airline-sentiment
 - Tweets.csv
- database.sqlite
 - Tweets

Output (56KB / 19.5GB)

/kaggle/working

Models

+ Add Models

No models added

Add a Kaggle model

Notebook options

- The independent variables of sentiment analysis are categorized.

TECHNIQUES FOR ANALYSIS:

BAG OF WORD[BoW]:

- The value in each dimension represents the frequency of the word's occurrence in the document.
- To perform sentiment analysis, you can use these word frequency vectors to train a machine learning model (e.g., Naive Bayes, Logistic Regression) to classify text into positive, negative, or neutral sentiments.

WORD EMBEDDINGS:

- Word embeddings capture semantic relationships between words. Words with similar meanings are represented as vectors close to each other in this vector space.
- In sentiment analysis, the document is represented as a combination of the word vectors within it.
- Then, the document representations are used as features to train machine learning models for sentiment classification.

Transformer Models: BERT (Bidirectional Encoder Representations from Transformers):

- BERT (Bidirectional Encoder Representations from Transformers) takes into account the context of a word by considering both the left and right context[captures the meaning of words in a more sophisticated manner].
- Transformers are known for their state-of-the-art performance in NLP tasks and have the ability to capture nuanced sentiment, context, and sarcasm in text.

Traditional Method for Sentiment Analysis:

- Using reference dictionaries , the traditional sentiment analysis assigns a score for each of the words in a sentence. The average of these scores is calculated and the sentiment of the text is then interpreted.

DEPLOYMENT OF MACHINE LEARNING:

- ML Ops is the process of deploying a machine learning model and integrating it into software that can be used by end users
- Machine learning involves giving data to a model to make predictions using statistical models, while deep learning uses artificial neural networks to process unstructured data like images, videos, and text.
- The common machine learning models:
 1. Logistic Regression
 2. Decision Trees
 3. Random Forest
 4. Support Vector Machines (SVM)
 5. k-NN
 6. Naive Bayes
 7. Neural Networks (Deep Learning)
 8. Clustering Algorithms
 9. Time Series Models

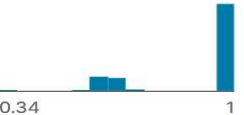
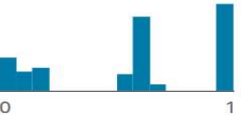
FEATURE EXTRACTION:

- We perform sentiment analysis using the independent variables and above ML models
- After we impute the dataset, the factors important for the sentiment in a response are filtered out.
- Consider, the tweet_id column in the dataset does not provide any form of sentiment. Likewise the values are noted down based on their weightage in the analysis.
- The dataset contains different reasons and texts, but the sentiments and sentiment confidence are groupable

← → ↻ kaggle.com/datasets/crowdfunder/twitter-airline-sentiment/data 🔍 📄 ⚙️ 🌐 🏠 📱

Tweets.csv (3.42 MB) 📄 ⚙️ <

Detail Compact Column 6 of 15 columns ▾

▲ airline_sentiment	# airline_sentiment...	▲ negativereason	# negativereason_c...	▲ airline	▲ text
negative 63%		[null] 37%		United 26%	14427 unique values
neutral 21%		Customer Service ... 20%		US Airways 20%	
Other (2363) 16%		Other (6268) 43%		Other (7905) 54%	
positive	0.3486		0.0	Virgin America	@VirginAmerica plus you've added commercials to the experience... tacky.
neutral	0.6837			Virgin America	@VirginAmerica I didn't today... Must mean I need to take another trip!
negative	1.0	Bad Flight	0.7033	Virgin America	@VirginAmerica it's really aggressive to blast obnoxious "entertainment" in your guests' faces &...
negative	1.0	Can't Tell	1.0	Virgin America	@VirginAmerica and it's a really big bad thing about it

VISUALIZATION AND INSIGHTS:

- After we analyze the dataset and create ML models, the data is visualized if necessary and the insights are generated/provided with reference to the data.
- In this project, we will use Artificial Neural Networks(ANN) model for the analysis of the dataset.
- After the successful development of the ML model, insights and visalization are obtained.
- The plotting of the dataset that has been analyzed is carried out using
 1. Pandas Plotting
 2. Matplotlib
 3. Seaborn