Sentiment Analysis For Marketing :

Introduction

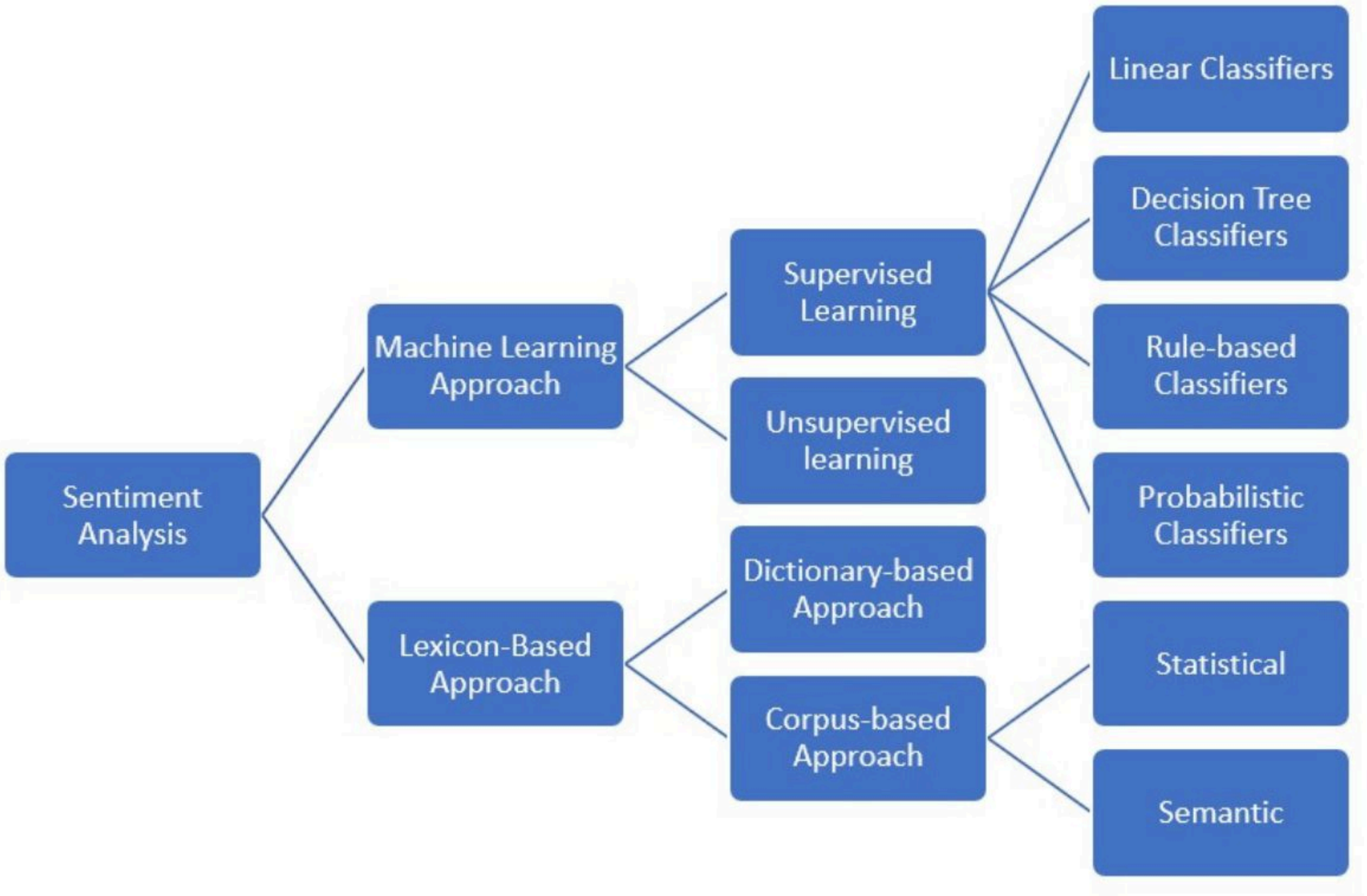
In recent years, consumers' willingness to share consumption experiences online, coupled with the technology to analyze "big data", o er marketing managers an unprecedented opportunity to collect market intelligence (Erevelles et al., 2016). Through online sentiment analysis, hereafter referred to as sentiment analysis, researchers can systematically extract and classify consumer emotions about products and services expressed in social network discussions and online postings to track brand attitudes and emerging market trends. While sentiment analysis presents tremendous opportunities to interpret a large body of data collected in a naturalistic setting, concerns have been expressed about the technique's accuracy and practicality (Gonçalves et al., 2013). Moreover, apprehensions over online data volume, fragmented data sources, content bias and user exploitation have exposed the technique to critical scrutiny.

Overview of sentiment analysis

While reviewing the literature, it is apparent that a misunderstanding often exists about what constitutes sentiment analysis. To provide conceptual clarity, sentiment analysis first needs to be distinguished from the broader literature on online text mining. With text-mining applications, researchers structure a large body of data from various online sources into numerous topics or themes which emerge from the body of textual data. In this regard, text mining is similar to traditional content analysis, since it allows researchers to e ciently extract, classify and manage a large body of data to identify hidden patterns or trends (He et al., 2013).

In contrast, sentiment analysis refers to the application of machine learning techniques to evaluate and classify attitudes and opinions on a specific topic of interest (Rambocas and Gama, 2013). Sentiment analysis focuses on extracting emotions from the online text but classifies specific problem areas into predefined mutually exclusive categories (Liu, 2012).

These categories imply bi-polar classifications of emotions (positive and negative) and are typically represented by numeric codes for subsequent statistical analyses.

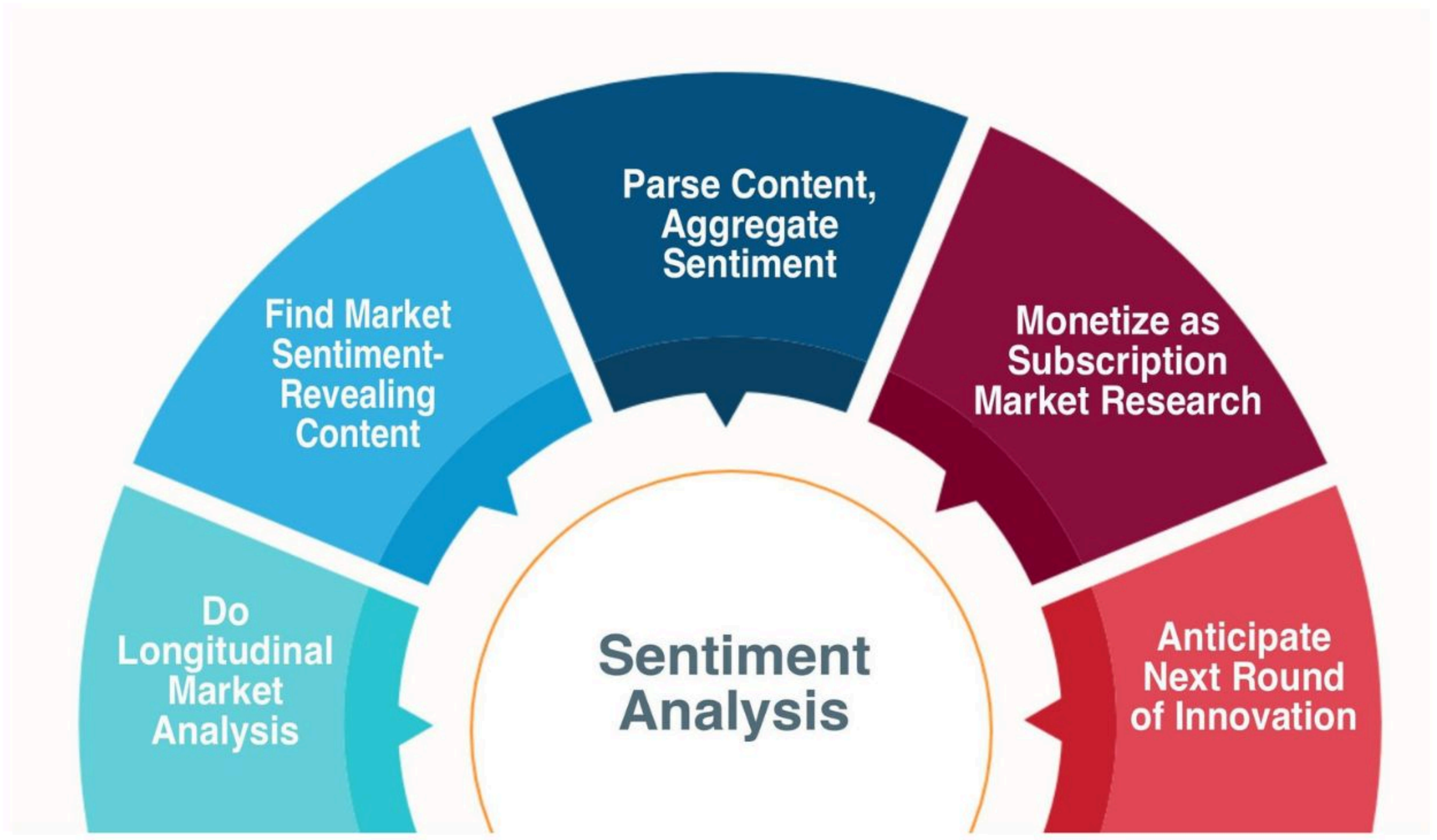


The scope and approach of the review

The objective of this review is to examine the use of sentiment analysis in the marketing literature published between 2008 and 2016. The studies identified in this paper were sourced using a combination of computerized and manual search methods. We first surveyed several online scientific databases including Business Source Complete, Proquest and Emerald, and conducted an issue by issue search of the top-ranked marketing journals. We also searched for articles using Google Scholar with the search term "sentiment analysis". Finally, we used a snowballing procedure where the references of each article on sentiment analysis were examined to identify additional studies, a search technique

consistent with Babic Rosario et al. (2016). This broad search yielded a total of 21,456 articles.

Next, the title, abstract and keywords of each article were examined to determine whether it was relevant to the application of online sentiment analysis or simply contained keywords such as "sentiments" or "emotions"



Characteristics of the marketing articles reviewed

The authors selected the 22 marketing articles and evaluated them on four criteria:

(1) utilized sentiment analysis and not text mining or other social media analytics;

(2) applied sentiment analysis in the study of marketing-related issue/s from the perspective of the consumer, business or both;
(3) published in a peer-reviewed academic journal;

(4) empirical in nature, with a large body of data and utilized statistical tests for data analysis.

Discipline	Frequency	(%)
Communication	26	10.12
Computer	185	71.98
Education	1	0.39
Engineering	2	0.78
Finance	5	1.95
Health	3	1.17
Marketing	22	8.56
Political Science	1	0.39
Review	12	4.67
Total	257	100

Methods used in sentiment detection and statistical analysis

Sentiment detection requires appraising and extracting only the emotionally laden content such as personal expressions, opinions and feelings from the textual data set. The studies reviewed in this paper employed either manual or automated detection mechanisms. Manual sentiment detection requires human input into the analysis and has the advantage of accommodating emoticons, abbreviations, sarcasm and slangs

Manual coding can also accommodate language idiosyncrasies. For instance, Liang et al. (2015) noted that the Part-of-Speech system in Taiwan is di erent from mainland China and English where many words that would be considered adjectives in both languages would be transitive verbs laden with sentiments. Although advantageous in many ways, manual coding can reflect individual subjectivity, bias and misinterpretations. Also, manual coding is very time-consuming and costly, with researchers spending several weeks coding and processing data into categories (Makarem and Jae, 2016).

The application of sentiment analysis in marketing research

In reviewing the application of sentiment analysis, we found the majority of articles focused on quantifying the e ect of online textual comments on corporate financial performance as measured by sales, preferential consumer behavior and corporate stock performance. In almost all cases, the research models were causal and driven by a strong theoretical underpinning.

For instance, Sonnier et al. (2011) considered the e ect of the volume of positive, negative and neutral user-generated comments on sales. The authors' study was among the first in the marketing literature to model the dynamic e ects of online communication and found positive feedback has the greatest e ect on sales followed by negative and neutral comments. In an extension of that work, Tang et al. (2014) showed that mixed-neutral comments intensify the impact of positive and negative comments while indi erent-neutral

PROGRAM

```
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python # For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
```

```
# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    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 output when you create a version using "Save & Run All"
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session
```

/kaggle/input/twitter-airline-sentiment/Tweets.csv

/kaggle/input/twitter-airline-sentiment/database.sqlite

In [2]:

```
import pandas as pd
df = pd.read_csv("/kaggle/input/twitter-airline-sentiment/Tweets.csv")
df
```

Out[2]:

	tweet_id	airline_sentiment	airline_sentiment_confidence	negativereason	negativereason_confidence	airline	airline_sentiment_gold	name	negativereason_gold	retweet_count	text	tweet_coord	tweet_created	tweet_location	user_timezone
0	570306133677760513	neutral	1.0000	NaN	NaN	Virgin America	NaN	cairdin	NaN	0	@VirginAmerica What @dhepburn said.	NaN	2015-02-24 11:35:52 -0800	NaN	Eastern Time (US & Canada)
1	570301130888	positive	0.3486	NaN	0.0000	Virgin America	NaN	jnardino	NaN	0	@VirginAmeric	NaN	2015-02-24	NaN	Pacific Time

	122368										a plus you've added commercials		11:15:59 -0800		(US & Canada)
2	570301083672 813571	neutral	0.6837	NaN	NaN	Virgin America	NaN	yvonnalynn	NaN	0	@VirginAmerica I didn't today... Must mean I n...	NaN	2015-02-24 11:15:48 -0800	Lets Play	Central Time (US & Canada)
3	570301031407 624196	negative	1.0000	Bad Flight	0.7033	Virgin America	NaN	jnardino	NaN	0	@VirginAmerica it's really aggressive to blast ...	NaN	2015-02-24 11:15:36 -0800	NaN	Pacific Time (US & Canada)
4	570300817074 4627 22	negative	1.0000	Can't Tell	1.0000	Virgin America	NaN	jnardino	NaN	0	@VirginAmeric a and it's a really big bad thing...	NaN	2015-02-24 11:14:45 -0800	NaN	Pacific Time (US & Canada)
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14635	569587686491 6825344	positive	0.3487	NaN	0.0000	American	NaN	KristenReenders	NaN	0	@AmericanAir thank you we got on a di erent f...	NaN	2015-02-22 12:01:01 -0800	NaN	NaN
14636	569587371693 355008	negative	1.0000	Customer Service Issue	1.0000	American	NaN	itsropes	NaN	0	@AmericanAir leaving over 20 minutes Late Flig...	NaN	2015-02-22 11:59:46 -0800	Texas	NaN
14637	569587242672 398336	neutral	1.0000	NaN	NaN	American	NaN	sanyabun	NaN	0	@AmericanAir Please bring American Airlines to...	NaN	2015-02-22 11:59:15 -0800	Nigeria ,lagos	NaN
14638	569587188687 634433	negative	1.0000	Customer Service Issue	0.6659	American	NaN	SraJackson	NaN	0	@AmericanAir you have my money, you change my ...	NaN	2015-02-22 11:59:02 -0800	New Jersey	Eastern Time (US & Canada)
14639	569587140490 866689	neutral	0.6771	NaN	0.0000	American	NaN	daviddtwu	NaN	0	@AmericanAir we have 8 ppl so we need 2 know h...	NaN	2015-02-22 11:58:51 -0800	dallas, TX	NaN

14640 rows × 15 columns

In [3]:

```
df.columns
```

Out[3]:

```
Index(['tweet_id', 'airline_sentiment', 'airline_sentiment_confidence',
       'negativereason', 'negativereason_confidence', 'airline',
       'airline_sentiment_gold', 'name', 'negativereason_gold',
       'retweet_count', 'text', 'tweet_coord', 'tweet_created',
       'tweet_location', 'user_timezone'],
      dtype='object')
```

In [4]:

```
df.isna().sum()
```

Out[4]:

```
tweet_id 0
airline_sentiment 0
airline_sentiment_confidence 0
negativereason 5462
negativereason_confidence 4118
```

airline 0

airline_sentiment_gold 14600

name 0

negativereason_gold 14608

retweet_count 0

text 0

tweet_coord 13621

tweet_created 0

tweet_location 4733

user_timezone 4820

dtype: int64

In [5]:

```
review_df = df[['text','airline_sentiment']]
```

```
print(review_df.shape)
```

```
review_df.head(5)
```

(14640, 2)

Out[5]:

	text	airline_sentiment
0	@VirginAmerica What @dhepburn said.	neutral
1	@VirginAmerica plus you've added commercials t...	positive
2	@VirginAmerica I didn't today... Must mean I n...	neutral
3	@VirginAmerica it's really aggressive to blast...	negative
4	@VirginAmerica and it's a	negative

	really big bad thing...	

In [6]:

```
print(review_df[3:4])
```

```
 text airline_sentiment

3 @VirginAmerica it's really aggressive to blast... negative
```

In [7]:

```
review_df = review_df[review_df['airline_sentiment'] != 'neutral']


print(review_df.shape)

review_df.head(5)
```

```
(11541, 2)
```

Out[7]:

	text	airline_sentiment
1	**@VirginAmerica plus you've added commercials t...**	**positive**
3	**@VirginAmerica it's really aggressive to blast...**	**negative**
4	**@VirginAmerica and it's a really big bad thing...**	**negative**
5	**@VirginAmerica seriously would pay $30 a fligh...**	**negative**
6	**@VirginAmerica yes, nearly every time I fly VX...**	**positive**

In [8]:

```
review_df["airline_sentiment"].value_counts()
```

Out[8]:

```
negative 9178

positive 2363

Name: airline_sentiment, dtype: int64
```

In [9]:

```
sentiment_label =

review_df.airline_sentiment.factorize()

sentiment _label
```

Out[9]:

```
(array([0, 1, 1, ..., 0, 1, 1]),

 Index(['positive', 'negative'], dtype='object'))
```

CONCLUSION

By understanding the sentiment of their customers, marketers can make better decisions about what to say and how to position their products or service. Additionally, sentiment analysis can help marketers identify potential issues with their product or service and address them before they become a problem.