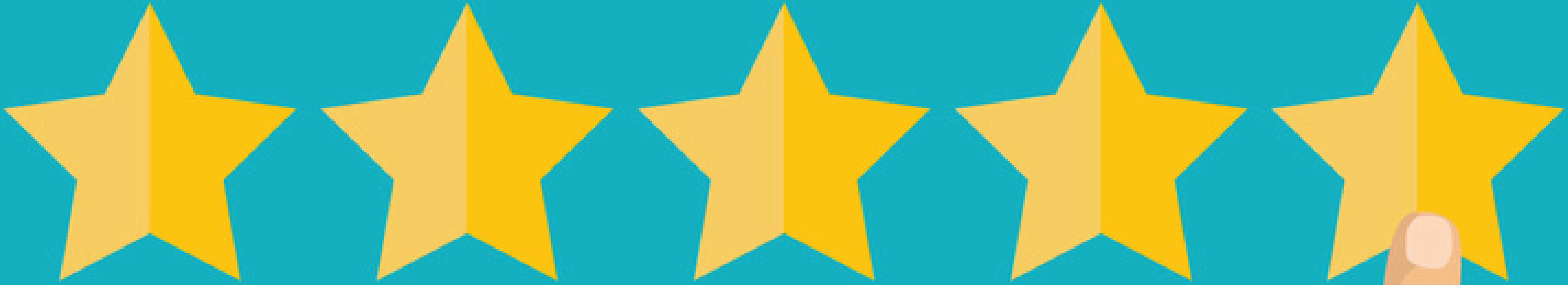
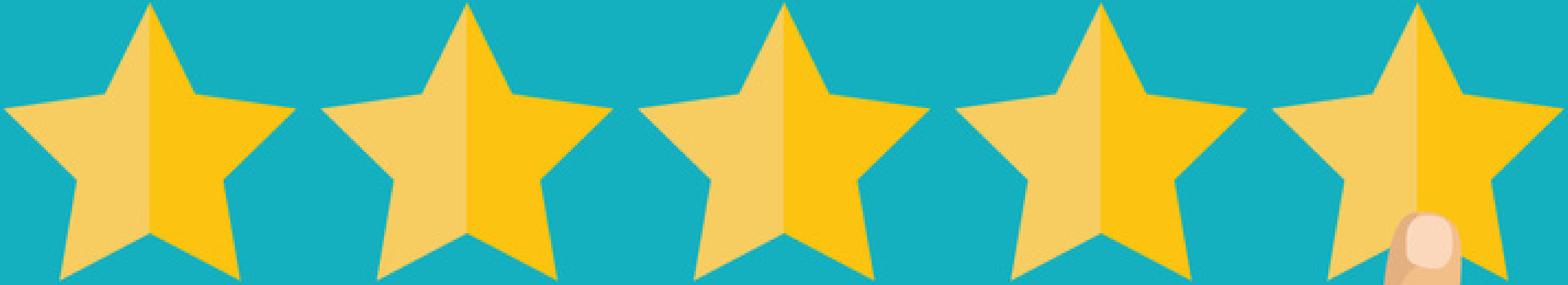


Amazon Reviews NLP Project

Category : Cell Phone & Accessories

Presented By
Pankaj V Tilwankar
PGA WE 07



SENTIMENT ANALYSIS OF E-COMMERCE PRODUCT REVIEWS

"Customer Voices, Amazon Choices: Your Reviews, Your Influence."

Presented By
Pankaj V Tilwankar
PGA WE 07

AGENDA

1. Introduction
2. Objective
3. What is Sentiment Analysis?
4. Why is Sentiment Analysis Important?
5. Process Flow
6. Challenges of Sentiment Analysis in E-commerce.

Introduction

- Have you ever looked at customer reviews before making a purchase?
- Of course you have! We all have. And that's where sentiment analysis comes in.
- Sentiment analysis is the process of analyzing customer reviews to determine the overall sentiment towards a product or service.
- It's becoming important for businesses to use sentiment analysis to improve their customer satisfaction and sales.
- But how does it work? That's what we're going to explore next.

What we are going to do?

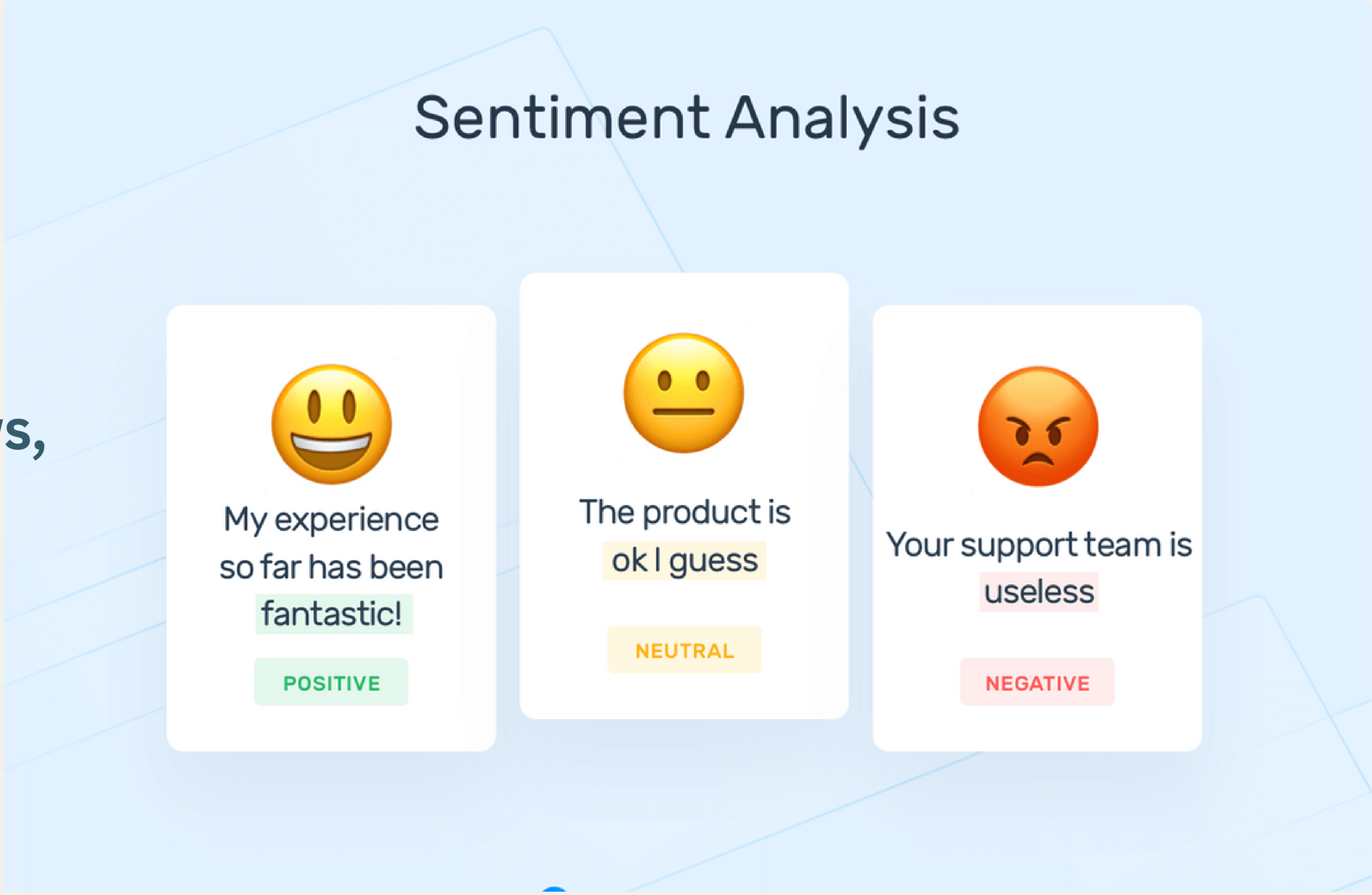
Develop an advanced Sentiment Analysis model using Natural Language Processing. (NLP) techniques and Machine Learning to accurately classify text data into sentiment categories, providing valuable insights for businesses and organizations to understand and respond to customer opinions effectively.

What is Sentiment Analysis?

- Sentiment analysis is the process of analyzing digital text to determine if the emotional tone of the message is positive, negative, or neutral.

- Also Know as Opinion Mining.

- This process is particularly useful in understanding the opinions, attitudes, and emotions conveyed by individuals or groups in various forms of communication, such as reviews, social media posts, news articles, and more.

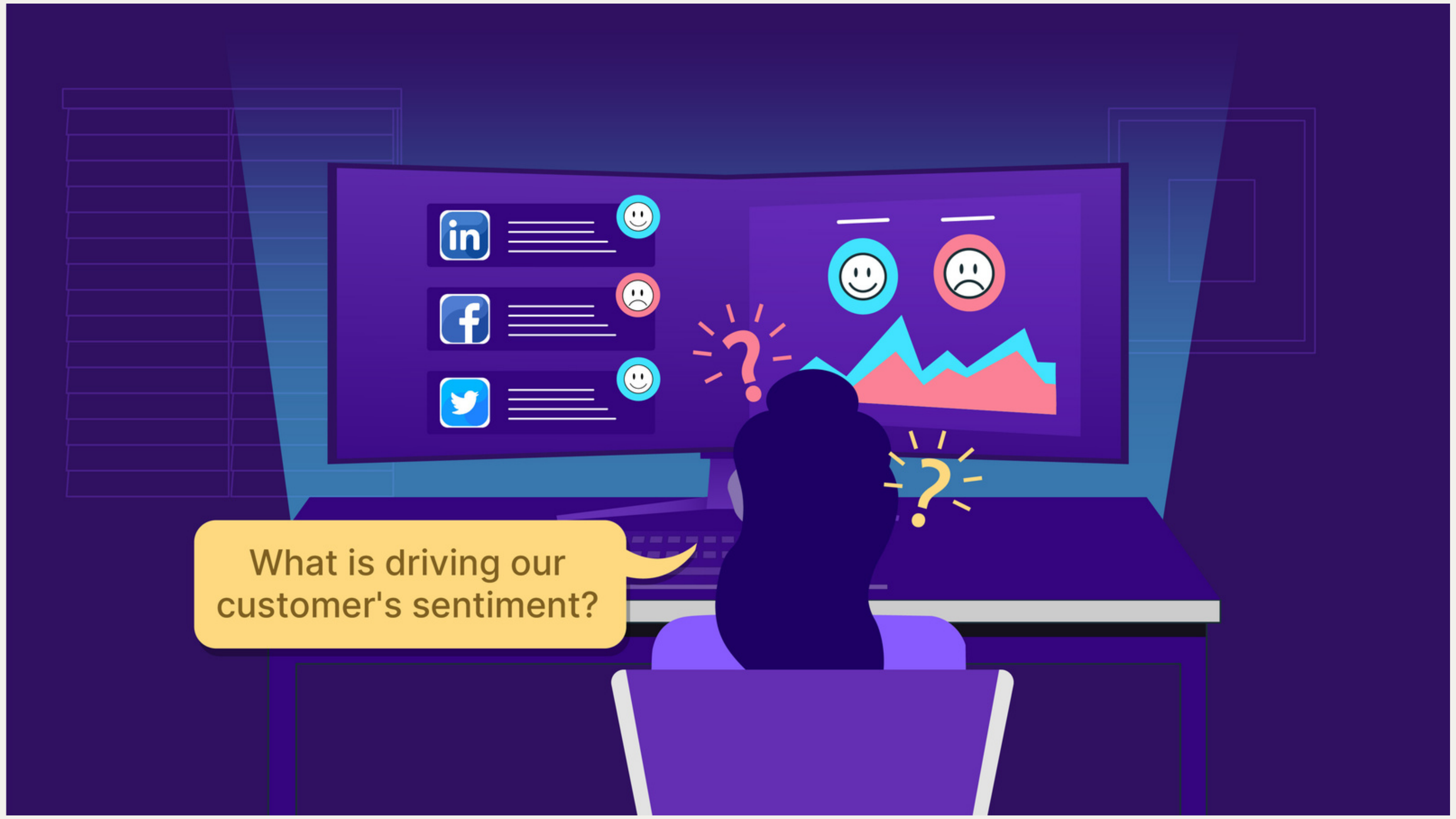


Why Sentiment Analysis is Important?

- Sentiment analysis allows businesses to understand the opinions, preferences, and emotions of their customers by analyzing reviews, feedback, and social media comments.

- It helps them understand customer feedback and improve their products and services accordingly.

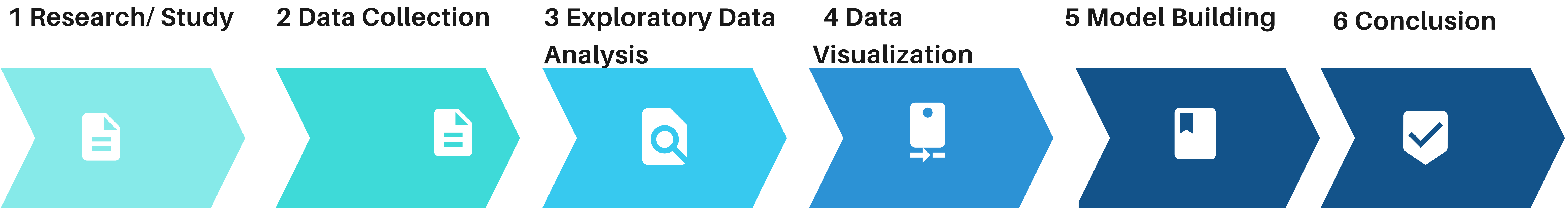
- Analyzing sentiment in customer support interactions helps companies identify areas for improvement in their service. It allows them to address customer concerns, improve response times, and enhance overall customer satisfaction.



Tool Used

- Platform / Tool : Jupyter Notebook
- Library Used : Scikit-learn, Pandas, Numpy
- Language : Python

PROCESS FLOW



Data Collection

Column Names :

1) reviewerID, 2) asin, 3) reviewerName, 4) helpful, 5) reviewText, 6) overall, 7) summary, 8) unixReviewTime, 9) reviewTime'

In [3]:

```
Data = pd.read_json(r"C:\Users\91983\Downloads\Amazon Review data set (Json)\Cell_Phones_and_Accessories_5.json" , lines = T
```

In [4]:

```
Data.columns
```

Out[4]:

```
Index(['reviewerID', 'asin', 'reviewerName', 'helpful', 'reviewText',
       'overall', 'summary', 'unixReviewTime', 'reviewTime'],
      dtype='object')
```

In [5]:

```
Data=Data.head(100000)
Data.head()
```

Out[5]:

	reviewerID	asin	reviewerName	helpful	reviewText	overall	summary	unixReviewTime	reviewTime
0	A30TL5EWN6DFXT	120401325X	christina	[0, 0]	They look good and stick good! I just don't li...	4	Looks Good	1400630400	05 21, 2014
1	ASY55RVNIL0UD	120401325X	emily l.	[0, 0]	These stickers work like the review says they ...	5	Really great product.	1389657600	01 14, 2014
2	A2TMXE2AFO7ONB	120401325X	Erica	[0, 0]	These are awesome and make my phone look so st...	5	LOVE LOVE LOVE	1403740800	06 26, 2014
3	AWJ0WZQYMYFQ4	120401325X	JM	[4, 4]	Item arrived in great time and was in perfect ...	4	Cute!	1382313600	10 21, 2013

Exploratory Data Analysis/ Data Cleaning

Steps :

1) Lower Case
2) Removing URL
3) Removing Punctuations
4) Removing Numbers
5) Removing Stopwords
6) Stemming/Lemmitization
7) Removing Extra white space

Term-Document Matrix is used to find the most important word in Dataset

Term-Document Matrix

In [20]:

```
from sklearn.feature_extraction.text import CountVectorizer
import pandas as pd

# Assuming df1 is your DataFrame with a column named 'reviewText' containing text data

# Create a CountVectorizer object and exclude common English stop words
cv = CountVectorizer(stop_words='english', max_features=5000)  # Limiting the number of features for illustration

# Fit and transform the 'reviewText' column to create a sparse document-term matrix
data_cv = cv.fit_transform(df.reviewText)

# Convert the sparse matrix to a DataFrame with feature names as columns
data_dtm = pd.DataFrame(data_cv.toarray(), columns=cv.get_feature_names_out())

# Transpose the DataFrame to create a term-document matrix
tdm = data_dtm.transpose()

# Display the first few rows of the term-document matrix
tdm.head()

```

Out[20]:

	0	1	2	3	4	5	6	7	8	9	...	99990	99991	99992	99993	99994	99995	99996	99997	99998	99999
abandoned	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
ability	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
able	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
abrasive	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
abroad	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

Measuring frequency of each word

In [21]:

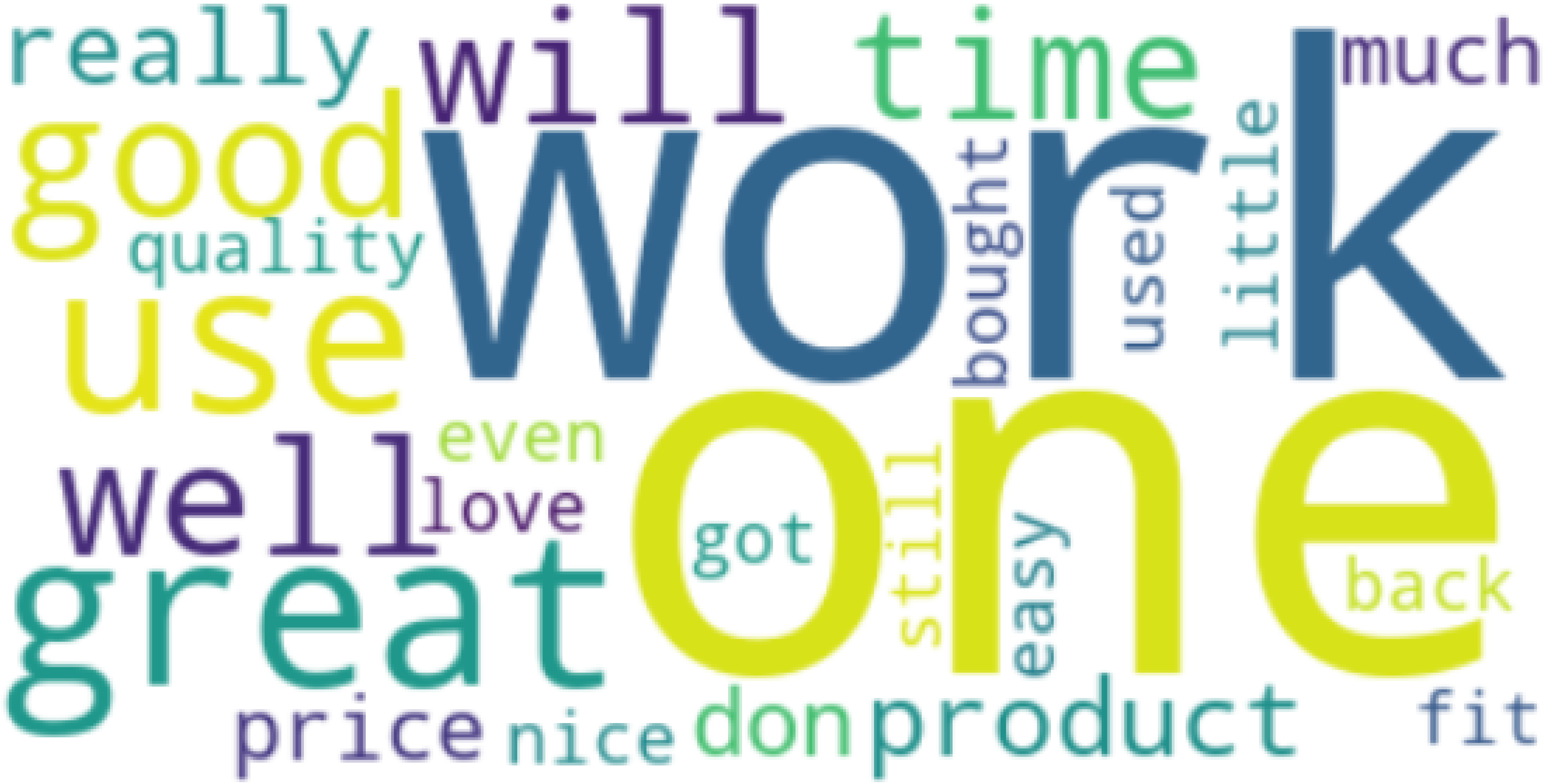
```
tdm['frequency'] = tdm.sum(axis=1)
tdm.head()
```

Out[21]:

	0	1	2	3	4	5	6	7	8	9	...	99991	99992	99993	99994	99995	99996	99997	99998	99999	frequency
abandoned	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	22
ability	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	938
able	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	5327
abrasive	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	22
abroad	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	53

5 rows × 100001 columns

Wordcloud


really
will
time
much
work
good
quality
use
bought
used
little
even
well
love
one
got
still
easy
great
back
price
nice
don
product
fit

Sentiment Analysis

- Using Textblob library each documents Polarity is calculated.
- Polarity range -1 to +1.
- -1 = Negative, +1 = Positive

Sentiment Analysis

In [31]:

```
from textblob import TextBlob

pol = lambda x: TextBlob(x).sentiment.polarity #polarity range -1 to +1
df['polarity'] = df['reviewText'].apply(pol)
df.head(5)
```

Out[31]:

	reviewText	polarity
0	They look good and stick good just don like th...	0.333333
1	These stickers work like the review they They ...	0.544444
2	These are awesome and make look stylish have o...	0.480000
3	Item great time and was perfect condition Howe...	0.600000
4	awesome stays and great can used multiple appl...	0.360000

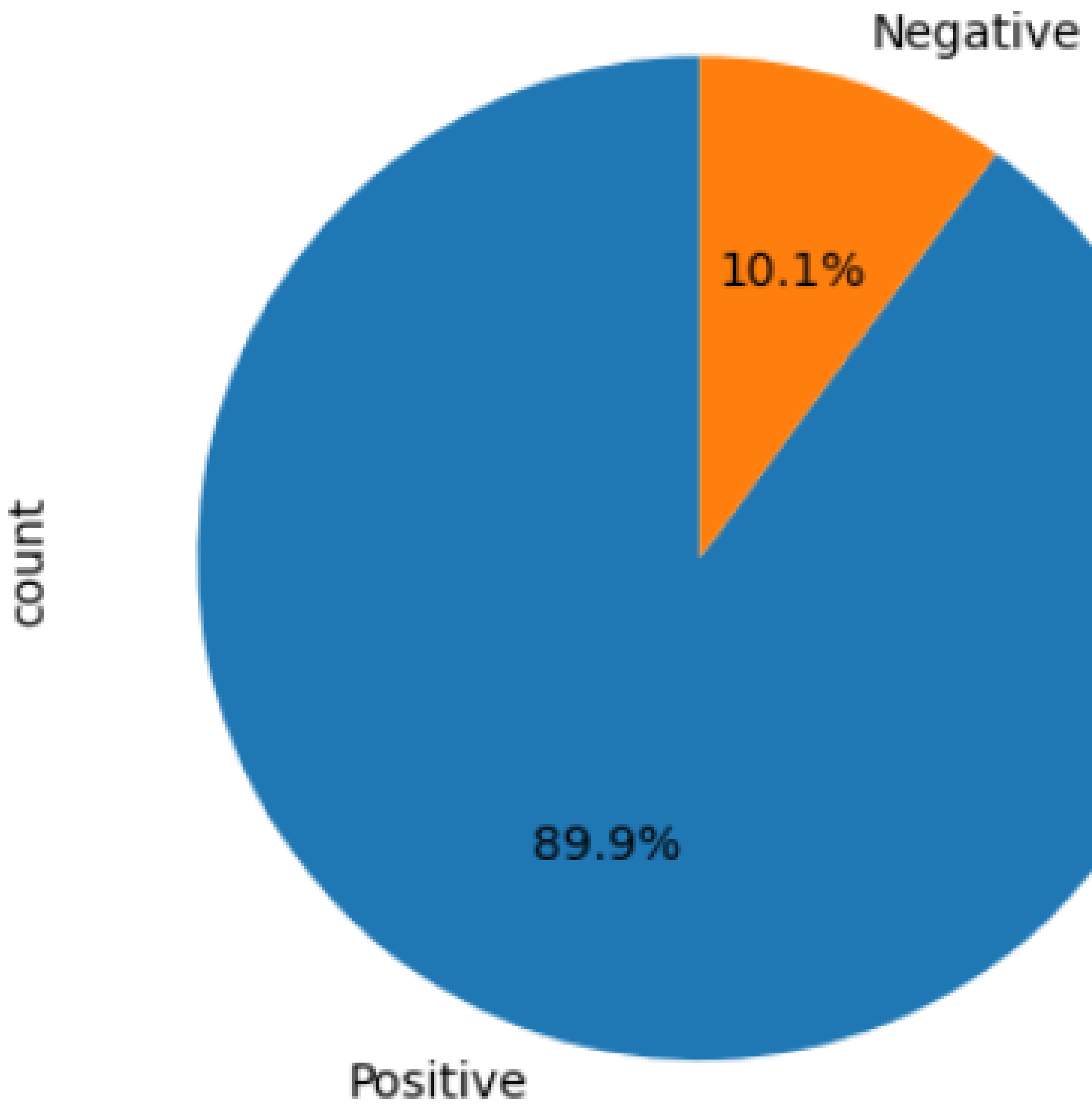
[32]:

```
import numpy as np
df['Sentiment'] = np.where(df['polarity']>= 0, 'Positive', 'Negative')
df.head()
```

[32]:

	reviewText	polarity	Sentiment
0	They look good and stick good just don like th...	0.333333	Positive
1	These stickers work like the review they They ...	0.544444	Positive
2	These are awesome and make look stylish have o...	0.480000	Positive
3	Item great time and was perfect condition Howe...	0.600000	Positive
4	awesome stays and great can used multiple appl...	0.360000	Positive

Graphical Representation of Sentiment



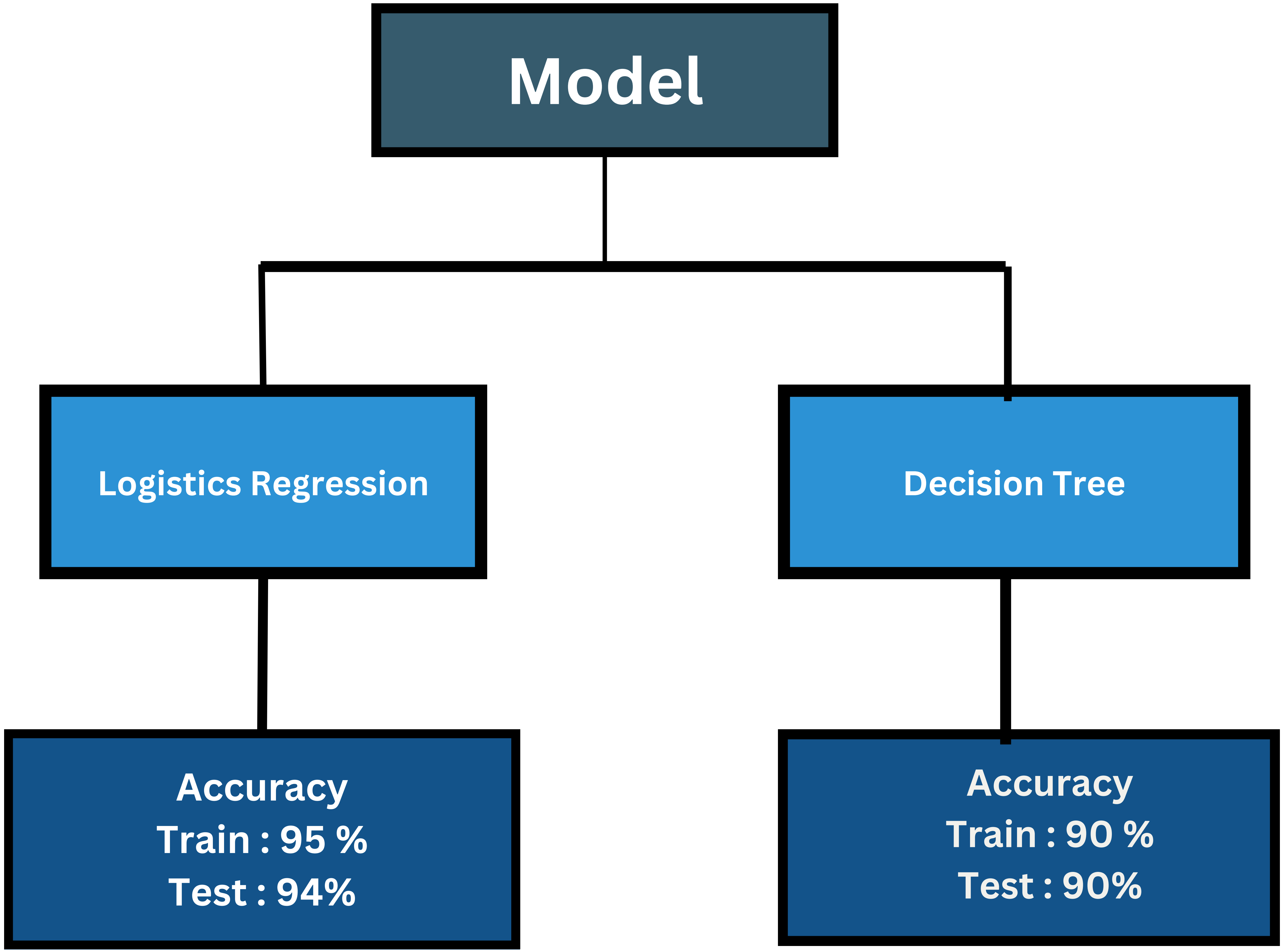
Model Building

Data Partition : Train data 70% , Test Data 30%

Data Partition

[39]:

```
#Dividing data into train and test dataset
from sklearn.model_selection import train_test_split
X = data_dtm.drop(['Sentiment'],axis=1)
Y = data_dtm['Sentiment']
X_train, X_test, y_train, y_test = train_test_split(X,Y,test_size=0.3,random_state=231)
```


Model
Logistics Regression
Decision Tree
Accuracy
Train : 95 %
Test : 94%
Accuracy
Train : 90 %
Test : 90%

MODEL SELECTION

Sr. No	Model	Accuracy(Train)	Accuracy(Test)
1	Logistics Regression	95 %	94 %
2	Decision Tree	90 %	90 %

- **Here we select Logistics Regression as a best model with high accuracy.**

Conclusion

- **Model Performance:** The accuracy of Model is 95%, demonstrating its effectiveness in classifying sentiment in customer reviews.
- **Key Features:** The analysis identified that certain keywords or phrases such as Good, Quality, Great, strongly influenced sentiment classification.
- **Real-world Application:** This sentiment analysis can be valuable to businesses for enhancing products, services and customer satisfaction by gaining insights from customer feedback.

Challenges of Sentiment Analysis

- **Ambiguity:** Sentences can have multiple meanings, making it challenging to determine the exact sentiment.
- **Negation:** Phrases like "not bad" can be tricky to interpret correctly, as they negate the sentiment.
- **Emojis and Emoticons:** Understanding the sentiment expressed through emojis and emoticons can be complex.
- **Short Texts:** Social media posts and text messages often contain very short and informal language, making it challenging to discern sentiment accurately.


Thank You