Introduction to twitter sentiment analysis

Twitter sentiment analysis involves examining the emotions and opinions expressed in tweets to understand public attitudes and trends. It provides valuable insights for businesses, marketers, and researchers.

Data Preprocessing

1. Text Cleaning

Removing hashtags, mentions, and URLs to focus on the actual text

```
# Defining regex patterns.
urlPattern        = r"((http://)[^ ]*|(https://)[^ ]*|( www\.)[^ ]*)"
userPattern       = '@[^\s]+'
alphaPattern      = "[^a-zA-Z0-9]"
sequencePattern   = r"(.)\1\1+"
seqReplacePattern = r"\1\1"
```

2. Tokenization

Breaking down tweets into individual words or tokens for analysis.

```
for word in tweet.split():
```

3. Stopword Removal

```
# Checking if the word is a stopword.
if word in stopwordlist:
    continue
#if word not in stopwordlist:
if len(word)>1:
    # Lemmatizing the word.
    word = wordLemm.lemmatize(word)
    tweetwords += (word+' ')
```

Training of model

Neural Networks

Utilizing deep learning techniques to understand complex patterns in data.

Dataset Preparation

Curating a diverse and representative dataset for effective training.

Performance Evaluation

Assessing the accuracy and effectiveness of the trained model.

Optimization

Tuning hyperparameters to enhance the model's performance.

Models used for training

1. BernoulliNB:

Type of Naive Bayes classifier, Effective for binary classification.

```
BNBmodel = BernoulliNB(alpha = 2)
BNBmodel.fit(X_train, y_train)
model_Evaluate(BNBmodel)
```

2. LinearSVC- Support Vector Classifier.

```
SVCmodel = LinearSVC()
SVCmodel.fit(X_train, y_train)
model_Evaluate(SVCmodel)
```

3. LR Model- Logistic Regression

Supervised learning algorithm for binary classification.

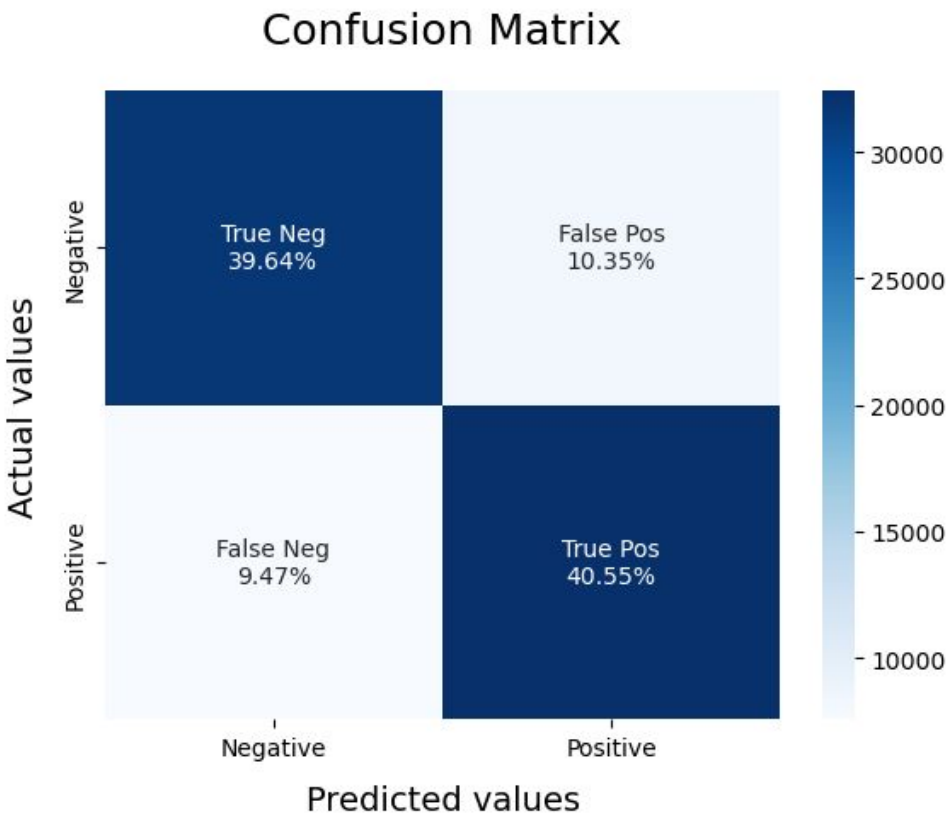
Uses Sigmoid function to find probability.

```
LRmodel = LogisticRegression(C = 2, max_iter = 1000, n_jobs=-1)
LRmodel.fit(X_train, y_train)
model_Evaluate(LRmodel)
```

Testing

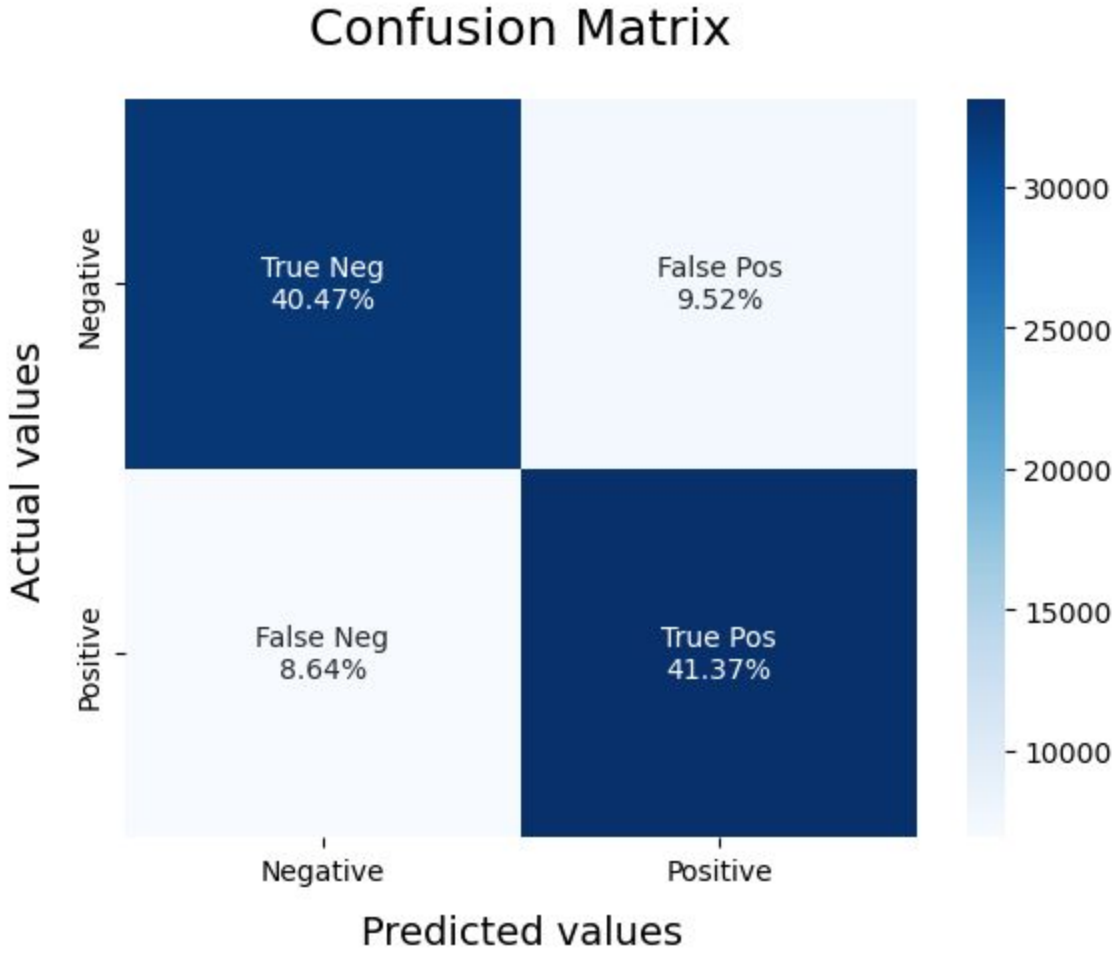
1- BernoulliNB

	precision	recall	f1-score	support
0	0.81	0.79	0.80	39989
1	0.80	0.81	0.80	40011
accuracy			0.80	80000
macro avg	0.80	0.80	0.80	80000
weighted avg	0.80	0.80	0.80	80000


Confusion Matrix
True Neg
39.64%
False Pos
10.35%
False Neg
9.47%
True Pos
40.55%
Negative
Positive
Actual values
Negative
Positive
Predicted values
30000
25000
20000
15000
10000

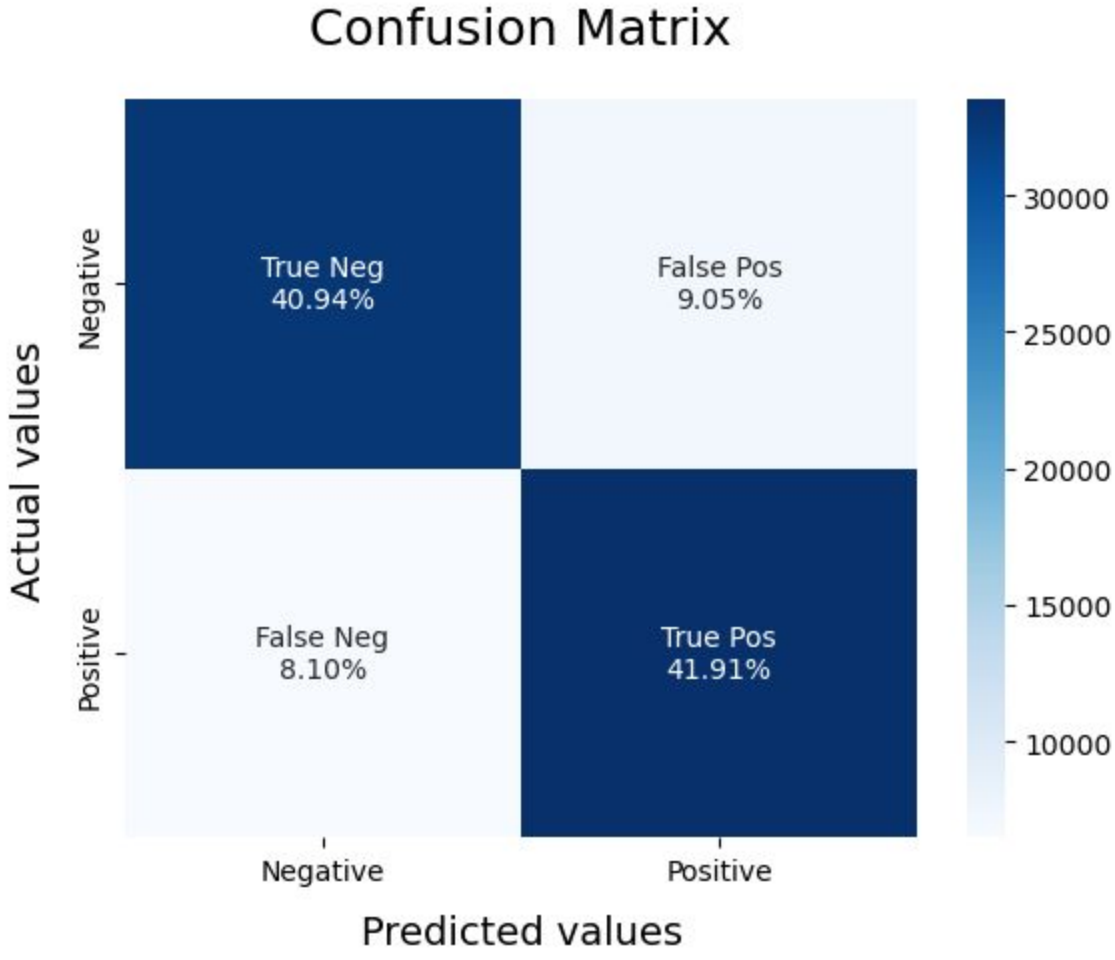
2- LinearSVC

	precision	recall	f1-score	support
0	0.82	0.81	0.82	39989
1	0.81	0.83	0.82	40011
accuracy			0.82	80000
macro avg	0.82	0.82	0.82	80000
weighted avg	0.82	0.82	0.82	80000


Confusion Matrix
True Neg
40.47%
False Pos
9.52%
False Neg
8.64%
True Pos
41.37%
Negative
Positive
Actual values
Negative
Positive
Predicted values
30000
25000
20000
15000
10000

3- LR Model

	precision	recall	f1-score	support
0	0.83	0.82	0.83	39989
1	0.82	0.84	0.83	40011
accuracy			0.83	80000
macro avg	0.83	0.83	0.83	80000
weighted avg	0.83	0.83	0.83	80000


Confusion Matrix
True Neg
40.94%
False Pos
9.05%
False Neg
8.10%
True Pos
41.91%
Negative
Positive
Actual values
Negative
Positive
Predicted values
30000
25000
20000
15000
10000

Conclusion and key takeaways

Insights Generation

Deriving actionable insights from sentiment analysis results to make informed decisions.

Real-Time Monitoring

Implementing continuous monitoring to stay updated on evolving trends.

Impactful Decision Making

Utilizing sentiment analysis to guide marketing strategies and brand perception.