

Natural Language processing Project

Sentiment analysis of Amazon Alexa reviews

Anudeep Yelakala

anudeepyelakala6@gmail.com

+917416915807

Links:

Google collab program file:

<https://colab.research.google.com/drive/1MRMwo9Inmr3i7IJkE8wUI2pnEXcsB2RC?usp=sharing>

Dataset drive link:

https://drive.google.com/file/d/1vVTf_fLugel_taweYogz8KY9unZuzzLK/view?usp=sharing

Problem statement

Amazon Alexa is a cloud-based voice service developed by Amazon that allows customers to interact with technology. There are currently over 40 million Alexa users around the world, so analyzing user sentiments about Alexa will be a good data science project. So, if you want to learn how to analyze the sentiments of users using Amazon Alexa, this article is for you. In this article, I'll walk you through the task of Amazon Alexa Reviews Sentiment Analysis Using Python.

Focuses to examine

- Data description and summary
- Exploratory data analysis
- Data Pre-processing (cleaning text, stop-words, vectorization)
- Machine learning algorithms
 1. Logistic Regression
 2. Multinomial naïve bayes
- Conclusion

Data description

The information contains data of various customer reviews of different amazon products

- Reviews: The verified reviews for the Alexa
- Feedback: 1(Positive) or 0 (Negative)

Visualizing data

```
[ ] dataset.head()
```

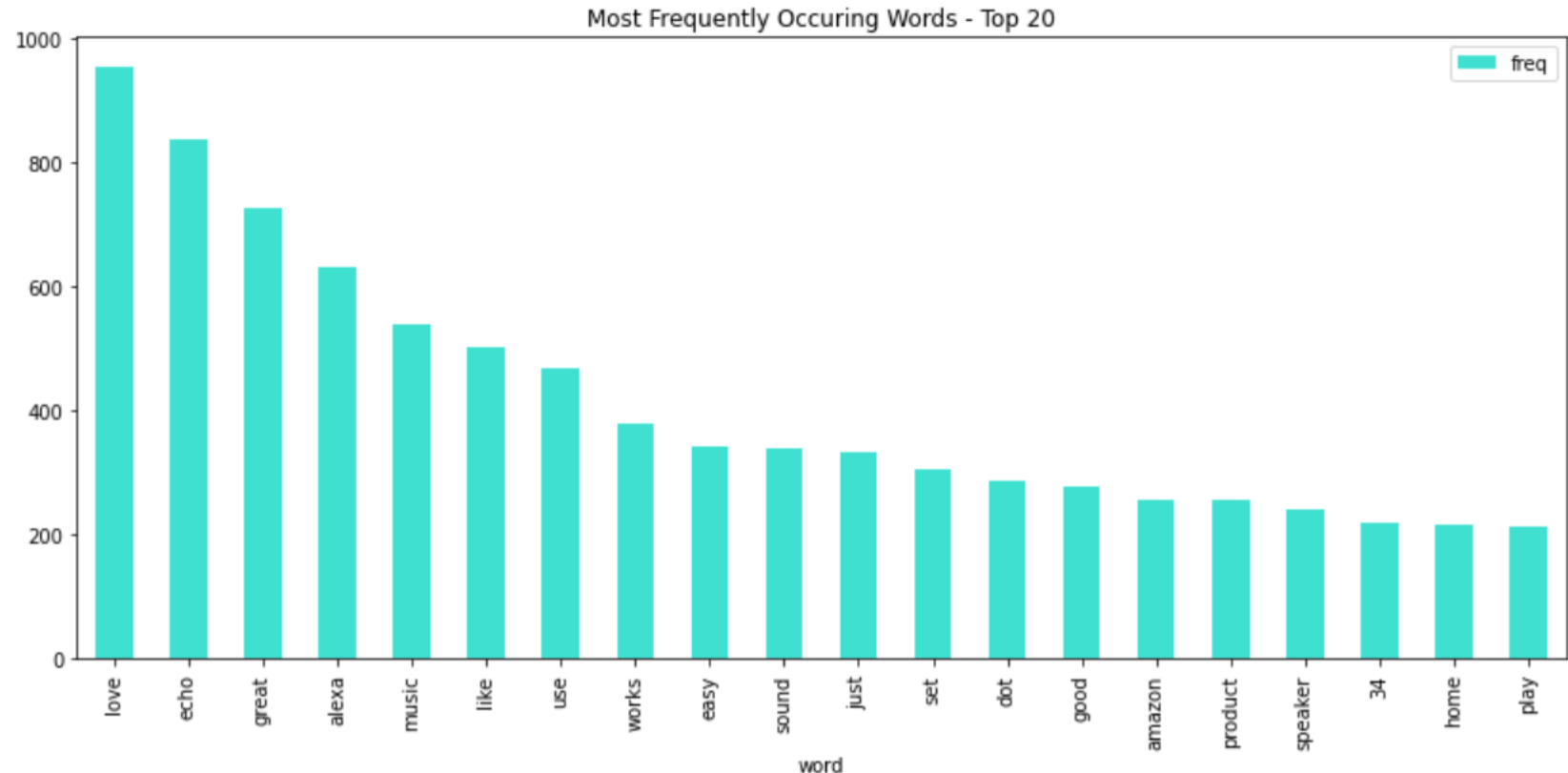
	reviews	sentiment
0	Love my Echo!	1
1	Loved it!	1
2	Sometimes while playing a game, you can answer...	1
3	I have had a lot of fun with this thing. My 4 ...	1
4	Music	1

```
dataset.tail()
```

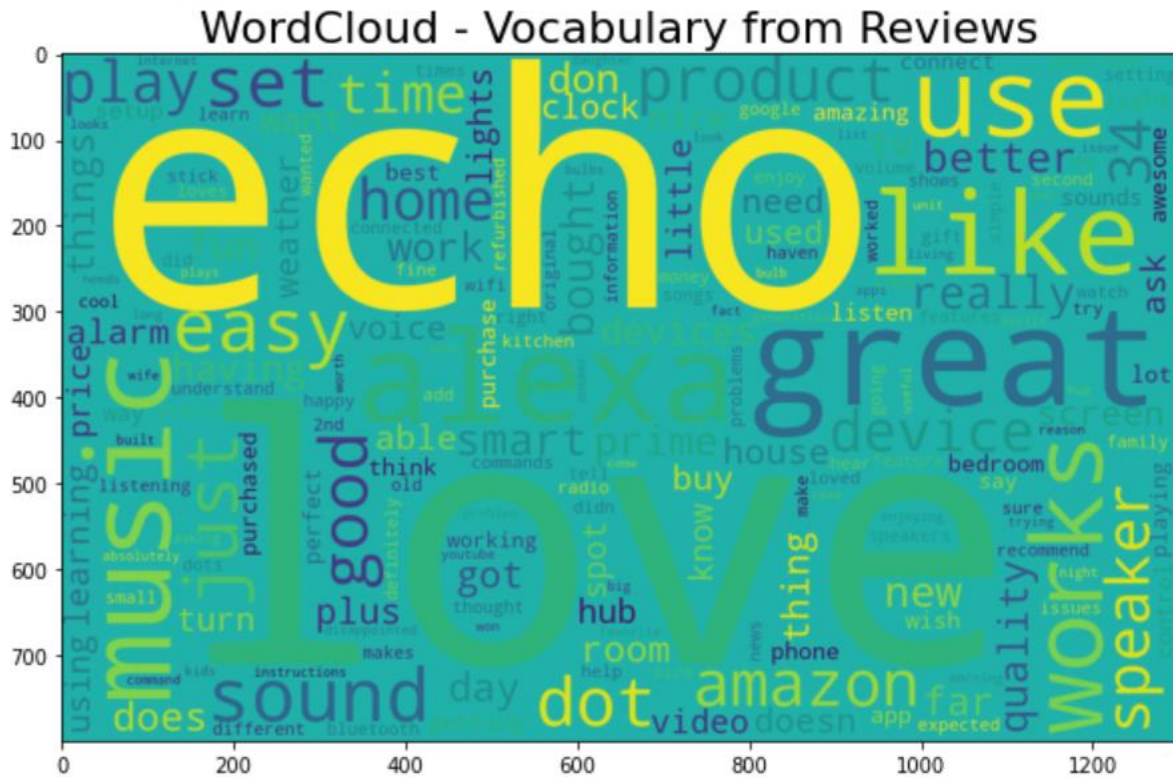
	reviews	sentiment
3145	Perfect for kids, adults and everyone in betwe...	1
3146	Listening to music, searching locations, check...	1
3147	I do love these things, i have them running my...	1
3148	Only complaint I have is that the sound qualit...	1
3149	Good	1

Most occurring words in the data

Text(0.5, 1.0, 'Most Frequently Occuring Words - Top 20')



Common words image



ML algorithms

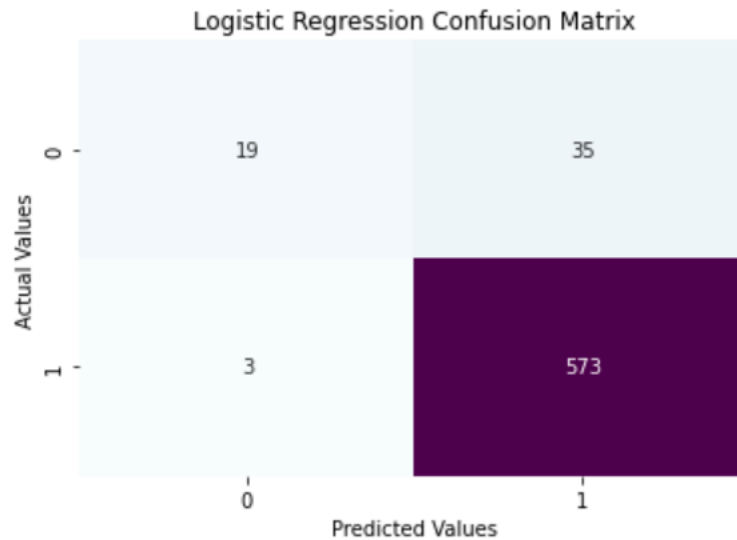
1. Logistic Regression
2. KNN Classifier
3. Multinomial Naive Bayes

- Logistic Regression

```
print('Logistic Regression Accuracy Score: ',accuracy_score(Y_test,y_pred))
cm=my_confusion_matrix(Y_test, y_pred, 'Logistic Regression Confusion Matrix')
```

Logistic Regression Accuracy Score: 0.9396825396825397

	precision	recall	f1-score	support
0	0.86	0.35	0.50	54
1	0.94	0.99	0.97	576
accuracy			0.94	630
macro avg	0.90	0.67	0.73	630
weighted avg	0.94	0.94	0.93	630

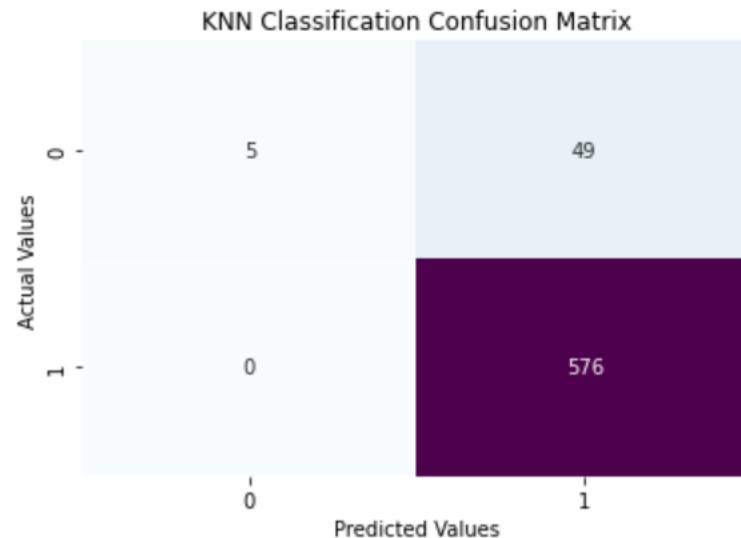


● KNN Classifier

```
print('KNN Classification Accuracy Score: ',accuracy_score(Y_test,y_pred))
cm=my_confusion_matrix(Y_test, y_pred, 'KNN Classification Confusion Matrix')
```

KNN Classification Accuracy Score: 0.9222222222222223

	precision	recall	f1-score	support
0	1.00	0.09	0.17	54
1	0.92	1.00	0.96	576
accuracy			0.92	630
macro avg	0.96	0.55	0.56	630
weighted avg	0.93	0.92	0.89	630

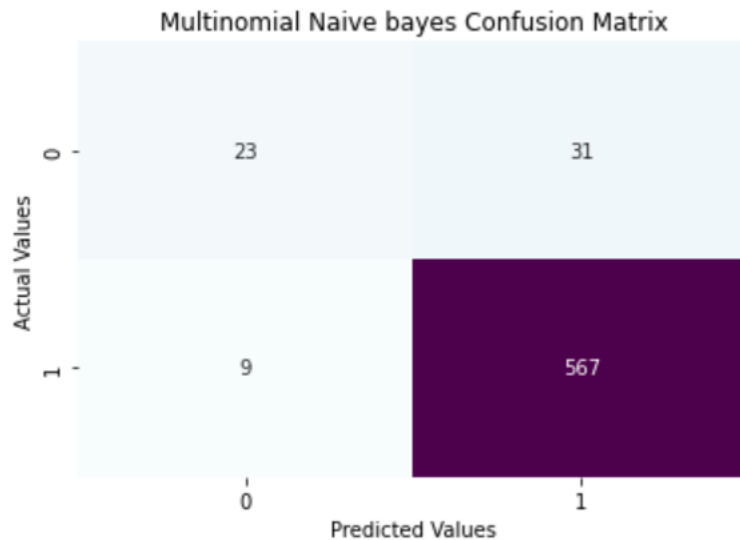


- Multinomial Naïve Bayes

```
print('Multinomial Navie Bayes Accuracy Score: ',accuracy_score(Y_test,y_pred))
cm=my_confusion_matrix(Y_test, y_pred, 'Multinomial Naive bayes Confusion Matrix')
```

Multinomial Navie Bayes Accuracy Score: 0.9365079365079365

	precision	recall	f1-score	support
0	0.72	0.43	0.53	54
1	0.95	0.98	0.97	576
accuracy			0.94	630
macro avg	0.83	0.71	0.75	630
weighted avg	0.93	0.94	0.93	630



conclusion

- Logistic Regression gave the highest accuracy score of 0.9396825396825397 followed by Multinomial bayes and KNN classification
- Therefore, most of the users feel neutral about the amazon products
- Almost 1936 users feel neutral about the products