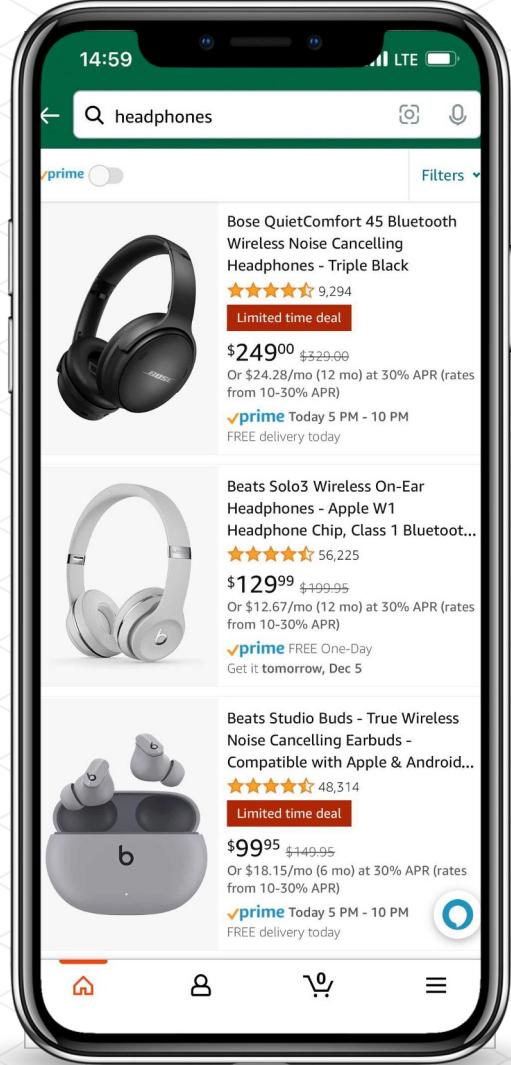




# Product Review Sentiment Analyzer based on **amazon**

Yanfeng(Jennifer) Zhang yz1045  
Zijing(Mavis) Cheng zc233  
Xin(Iris) Xiang xx123



1

Brand ?

2

Price ?

3

Product Information ?



Sony WH-1000XM4  
Wireless Premium Noise  
Canceling Overhead He...



\$348<sup>00</sup>

✓prime



Sony WH-1000XM5  
Wireless Industry Leading  
Noise Canceling Headp...



\$348<sup>00</sup>

✓prime

# ⭐ Rating ⭐

## Customer reviews

★★★★★ 4.7 out of 5

44,303 global ratings

[Write a review](#)

5 star

4 star

3 star

2 star

1 star

99%

9%

3%

2%

3%

FAKE

# ⭐ Rating ⭐

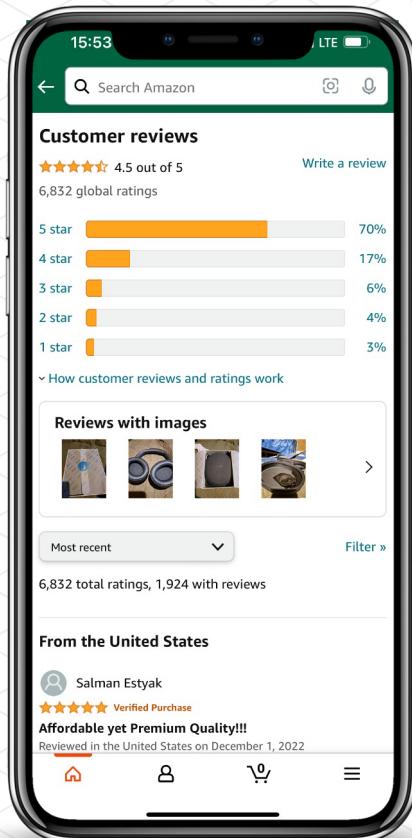
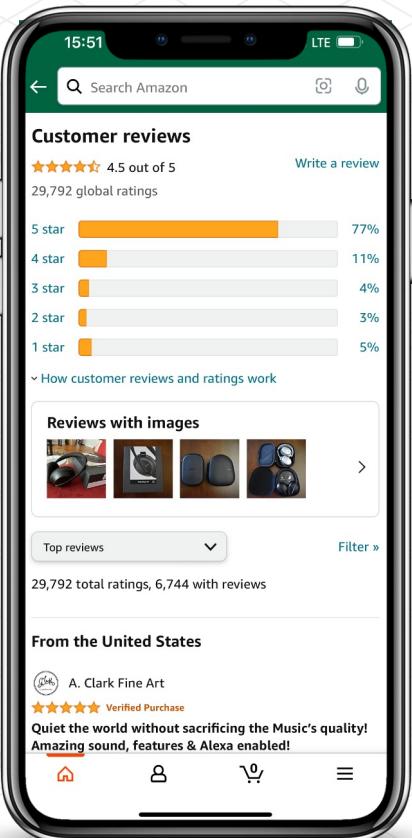


-- Positive

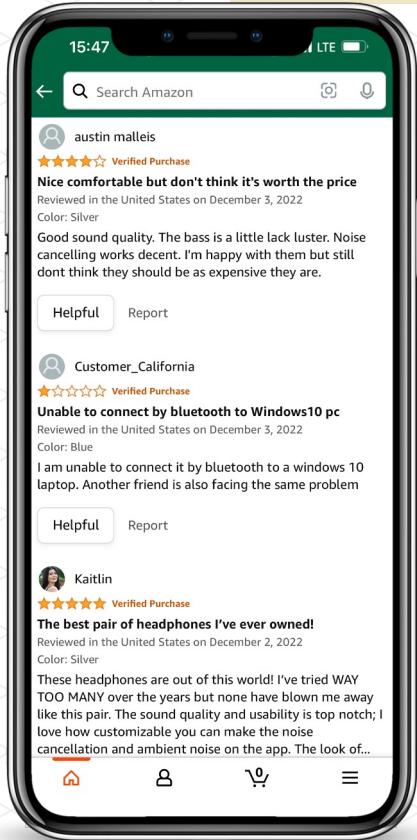
-- ?

-- Negative

# ⭐ Rating ⭐



# Review



**Customer  
Rating**



**Customer  
Review**

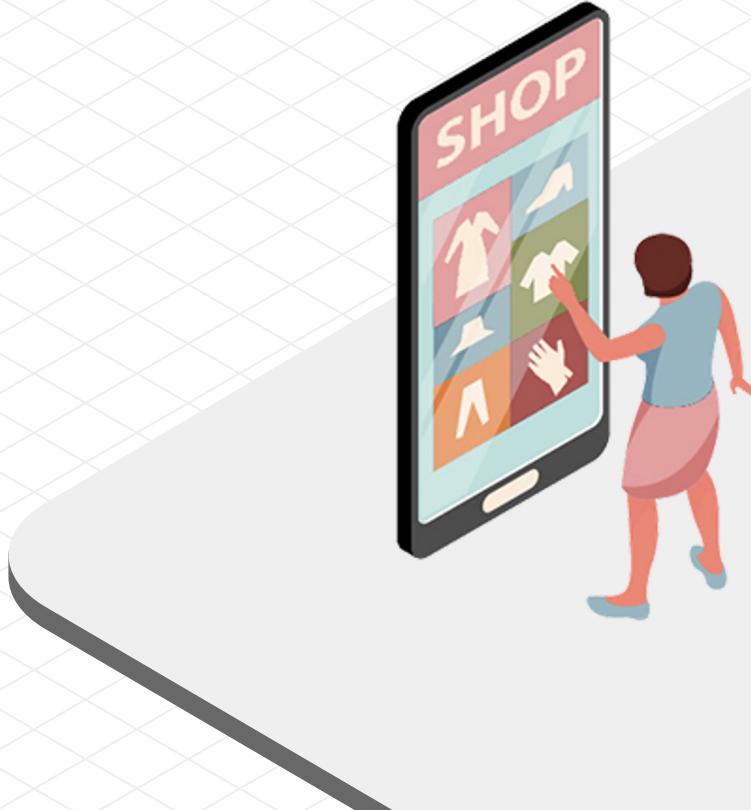
**Customer  
Rating**

**VS**

**Customer  
Review**

# Our Project

- ✓ Timesaving
- ✓ Intuitive
- ✓ Remove Bias:  
Fake Reviews
- ✓ Relatively Fair



01

Data Collection

02

Data  
Preprocessing

03

Methodology

04

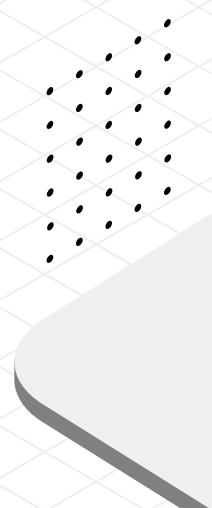
Model  
Evaluation &  
Model  
Selection

05

Outcomes

06

Conclusion



01

# Data Collection



# Data Introduction



Resource

150 Products

103008 Reviews



Category

# Data Collection

SORT BY

FILTER BY

Top reviews

All reviewers

All stars

All formats

Text, image, video

20,215 total ratings, 2,946 with reviews

## From the United States



Thomas P. Panfile

★★★★★ Broke after one use! (See update)  
Reviewed in the United States on November 8, 2022

Color: Blue | Style: WHCH710N

Verified Purchase

Noise canceling is 3-4 stars. Sound quality is 4 stars. Build quality is 1 star. The foam ear cushion came off after one use by my wife. I was able to get it back on. But I'm not sure if it will stay on. I also own the \$348 Sony model WH-1000XM4 which I use. After 2 years and very heavy daily use, this model still works great. Maybe her unit was a lemon, but I'd get the \$348 model.

Update:

My wife called Sony and was immediately connected to a customer service rep. They are sending her new headphones! Very good response...."Sony No Baloney"! Upgraded from 2 stars to 4 stars

71 people found this helpful

Helpful

| Report abuse

# Data Collection



<Requests>

	product_name	username	verification	point	review	detail	helpful
0	Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	ShepardsFool	Verified Purchase	1.0 out of 5 stars	<span>Died in 2 days. Do not buy</span>	<span>Within 2 days the right ear bud stopped ...	
1	Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	jerry	Verified Purchase	1.0 out of 5 stars	<span>Essentially worthless</span>	<span>I ordered these and another pair from 1M...	
2	Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	J. Barlow	Verified Purchase	5.0 out of 5 stars	<span>Great sound</span>	<span>These little buds pick up music greatly....	
3	Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	lbt	Verified Purchase	2.0 out of 5 stars	<span>Poor sound</span>	<span>I don't know that \$8 is worth returning....	
4	Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	Katerine Jahne	Verified Purchase	5.0 out of 5 stars	<span>Work Great!</span>	<span>What can you say, they're headphones. Wo...	
...	...	...	...	...	...	...	...
103003	Skullcandy Hesh Evo Wireless Over-Ear Headphon...	Jayakrishna B.	Verified Purchase	5.0 out of 5 stars		<span>Happy with the product</span>	
103004	Skullcandy Hesh Evo Wireless Over-Ear Headphon...	Aamitesh Banerjee	Verified Purchase	5.0 out of 5 stars		<span>So, this one is actually a 5 on 5 for me...	
103005	Skullcandy Hesh Evo Wireless Over-Ear Headphon...	Anand	Verified Purchase	5.0 out of 5 stars		<span>If you're looking for a great sounding w...	<span class="cr-original-review-content" data-hook="helpful-vote-statement">&lt;span&gt;Mic quality is very poor , sound and noi...</span>
103006	Skullcandy Hesh Evo Wireless Over-Ear Headphon...	julio alberto	Verified Purchase	5.0 out of 5 stars		<span>class="cr-original-review-content">sin c...	
103007	Skullcandy Hesh Evo Wireless Over-Ear Headphon...	Navin	Verified Purchase	3.0 out of 5 stars			

103008 rows x 7 columns

02

# Data Preprocessing



# Data Preprocessing

## Data Cleaning

1. Non-relative Variables
2. Missing Values
3. Punctuation
4. Emojis



product_name	username	verification	point	date	review	detail
Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	ShepardsFool	Verified Purchase	1.0 out of 5 stars	Reviewed in the United States on November 22, ...	<span>Died in 2 days. Do not buy</span>	<span>Within 2 days the right ear bud stopped ...
Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	jerry	Verified Purchase	1.0 out of 5 stars	Reviewed in the United States on November 22, ...	<span>Essentially worthless</span>	<span>I ordered these and another pair from 1More. T...
Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	J. Barlow	Verified Purchase	5.0 out of 5 stars	Reviewed in the United States on November 22, ...	<span>Great sound</span>	<span>These little buds pick up music greatly....
Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	lbt	Verified Purchase	2.0 out of 5 stars	Reviewed in the United States on November 22, ...	<span>Poor sound</span>	<span>I don't know that \$8 is worth returning,...
Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	Katerine Jahne	Verified Purchase	5.0 out of 5 stars	Reviewed in the United States on November 21, ...	<span>Work Great!</span>	<span>What can you say, they're headphones. Wo...



product_name	username	verification	point	review	detail
Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	ShepardsFool	Verified Purchase	1.0	Died in 2 days. Do not buy	Within 2 days the right ear bud stopped workin...
Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	jerry	Verified Purchase	1.0	Essentially worthless	I ordered these and another pair from 1More. T...
Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	J. Barlow	Verified Purchase	5.0	Great sound	These little buds pick up music greatly. I ca...
Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	lbt	Verified Purchase	2.0	Poor sound	I don't know that \$8 is worth returning, but u...
Panasonic ErgoFit Wired Earbuds, In-Ear Headph...	Katerine Jahne	Verified Purchase	5.0	Work Great!	What can you say, they're headphones. Work well.

# Data Collection

SORT BY

FILTER BY

Top reviews

All reviewers

All stars

All formats

Text, image, video

20,215 total ratings, 2,946 with reviews

## From the United States



Thomas P. Panfile

★★★★★ Broke after one use! (See update)  
Reviewed in the United States on November 8, 2022

Color: Blue | Style: WHCH710N

Verified Purchase

Noise canceling is 3-4 stars. Sound quality is 4 stars. Build quality is 1 star. The foam ear cushion came off after one use by my wife. I was able to get it back on. But I'm not sure if it will stay on. I also own the \$348 Sony model WH-1000XM4 which I use. After 2 years and very heavy daily use, this model still works great. Maybe her unit was a lemon, but I'd get the \$348 model.

Update:

My wife called Sony and was immediately connected to a customer service rep. They are sending her new headphones! Very good response...."Sony No Baloney"! Upgraded from 2 stars to 4 stars

71 people found this helpful

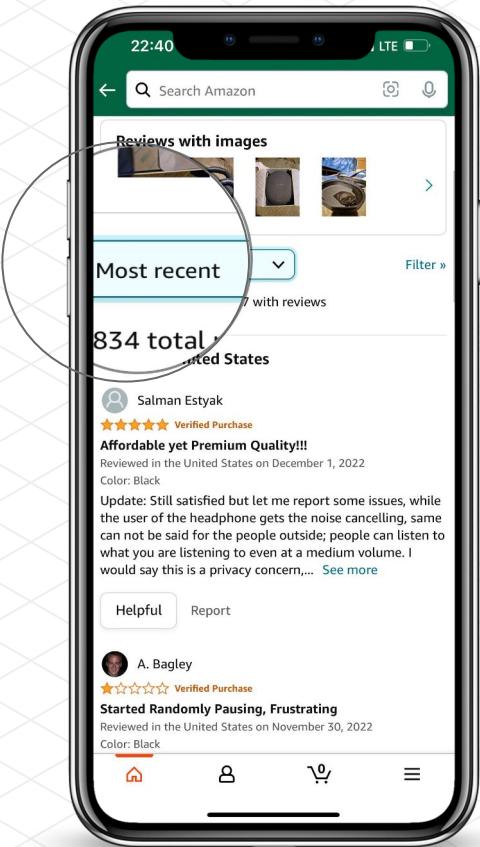
Helpful

| Report abuse

# Data Preprocessing

## Remove Fake Review

1. "Most Recent"
2. Review Length
3. Duplication
4. Similarity



# Data Preprocessing

## Remove Fake Review

1. "Most Recent"
2. Review Length
3. Duplication
4. Similarity



🚫 Length less than 5

🚫 Reviews from same user  
within 7 days

🚫 Review repetition rate  
large than 98%

# Data Preprocessing

Tokenization



Vectorization

TfidfTransformer()





03

# Methodology

# General ideas

## Vectorization

We build a feature vector that can be used for testing the models



## Machine Learning Methods and Deep Learning Methods

We divide training into two parts, machine learning methods and deep learning methods. In machine learning part, we choose SVM, SGD, RF, NB and LR while in deep learning part, we choose DNN(Dense Neural Network) model, LSTM, BiLSTM and Conv1D.



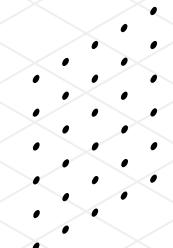
## Parameter Tuning

The best performance in these models is SVC with polynomial kernel



## Prediction on testing set

Finally, we predict the testing set with the best model



# Sentiment Labeling

★★★★★ Verified Purchase

## Good Noise Cancellation

Reviewed in the United States on November 29, 2022

Color: Black

The good:

- Noise Cancellation works pretty well
- Foldable, sturdy pouch
- Good battery life and fast charging

The so and so:

- Really bad thermal noise when using with the audio jack (on laptop)
- Tricky to use noise cancellation with the audio jack
- Can't charge and use wirelessly
- really bad voice isolation (to be fair this has never been part of the features, i rechecked), yet if you are looking for a headset that would isolate your voice from your surroundings when taking a call, this is not for you.

Helpful

Report

★★★★★

**Great sound, cheap build. Will break in a year :(**

Reviewed in the United States on September 20, 2022

Color: Pink



(9/20/2022) I absolutely love these headphones! They sound great, are easy to use, and have a nice app. However, I've had them for about a year now and they're completely falling apart. The plastic surrounding the crews connecting the earphone to the headband is

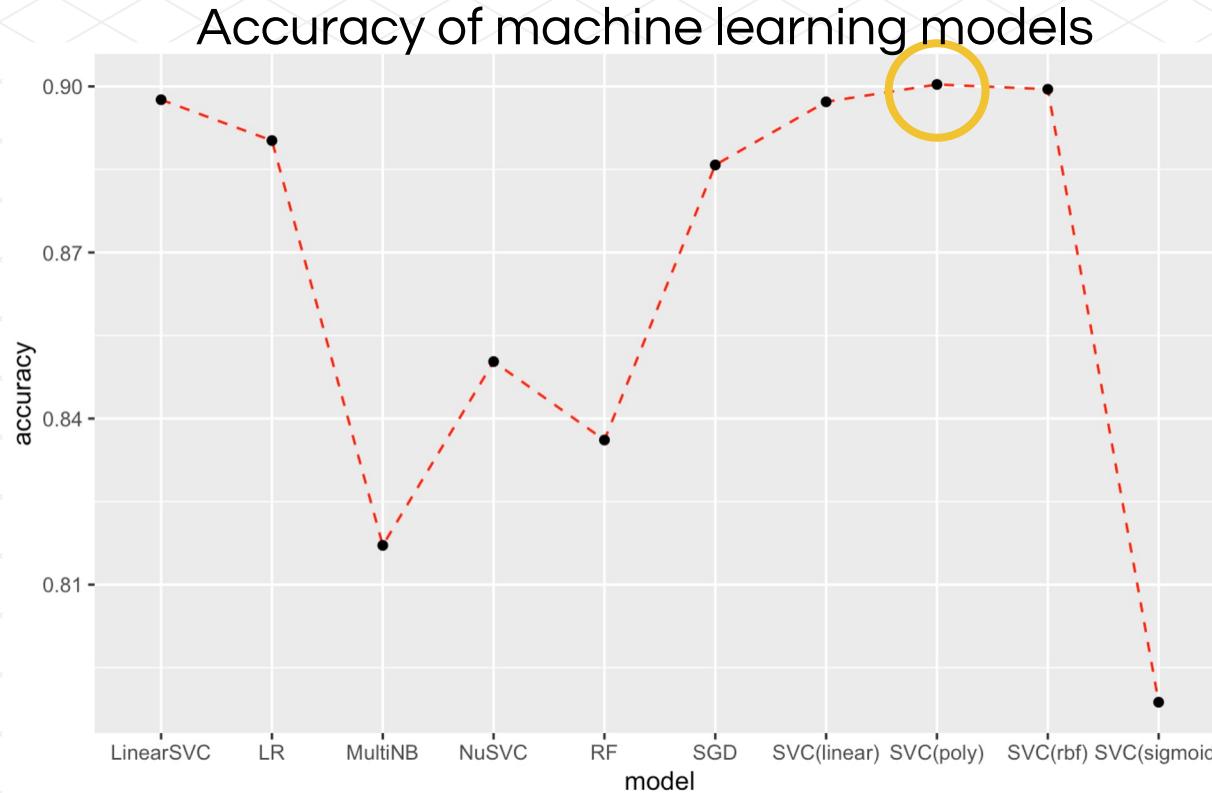


A 3D-style illustration of a woman with dark hair tied back, wearing a pink long-sleeved shirt and blue jeans, pushing a shopping cart. The cart is filled with various grocery items like a red bag, a water bottle, a box of cereal, and a bottle of milk. The scene is set on a light brown surface, which appears to be a smartphone screen, as indicated by the grid pattern and rounded corners.

04

# Model Evaluation and Model Selection

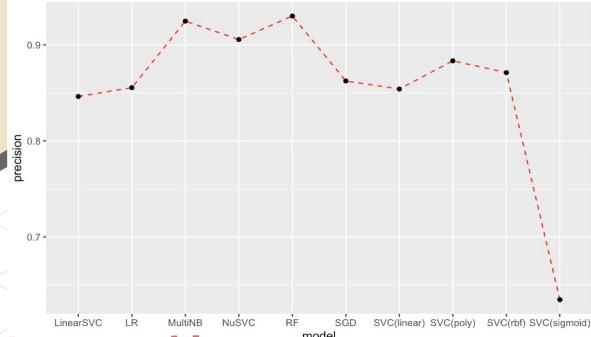
# Model evaluation-Machine Learning Models



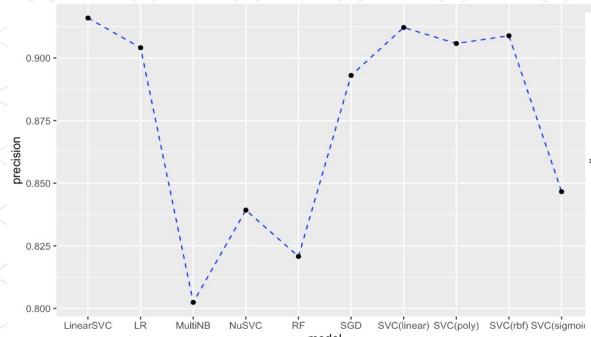
Based on accuracy,  
**SVC with polynomial kernel**  
performs the best on validation set

# Model evaluation-Machine Learning Models

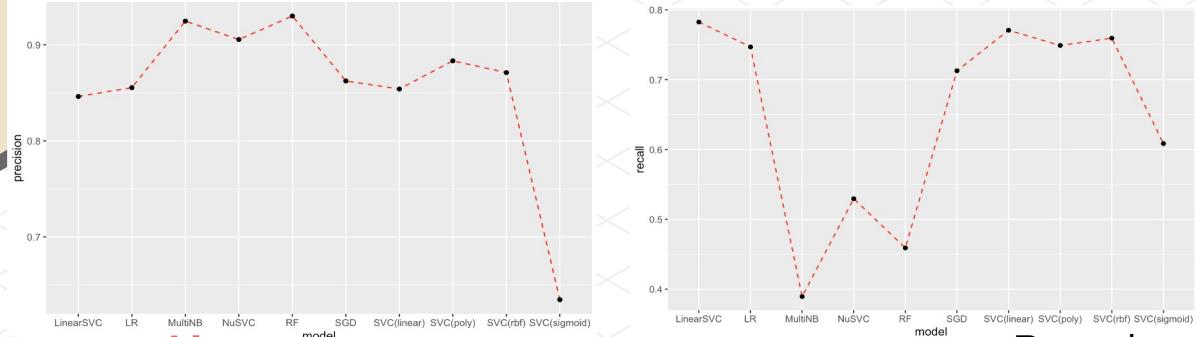
Precision of machine learning models Recall of machine learning models F1-score of machine learning models



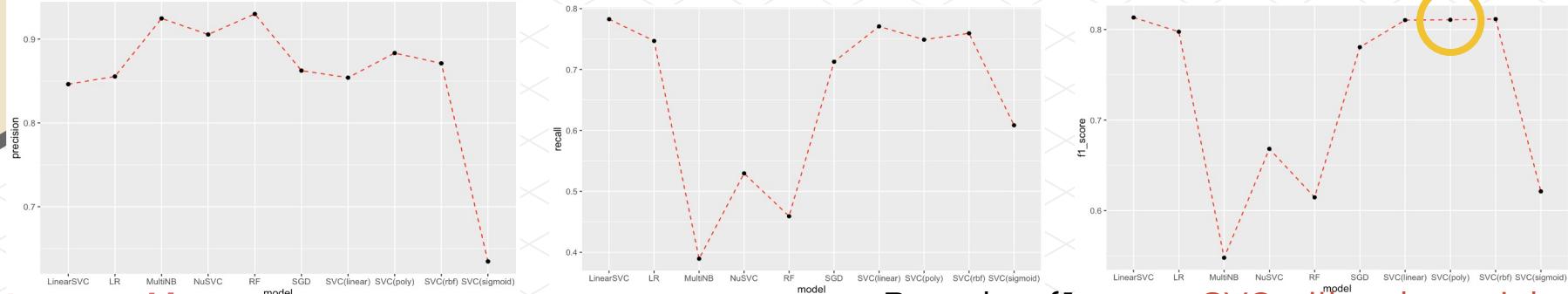
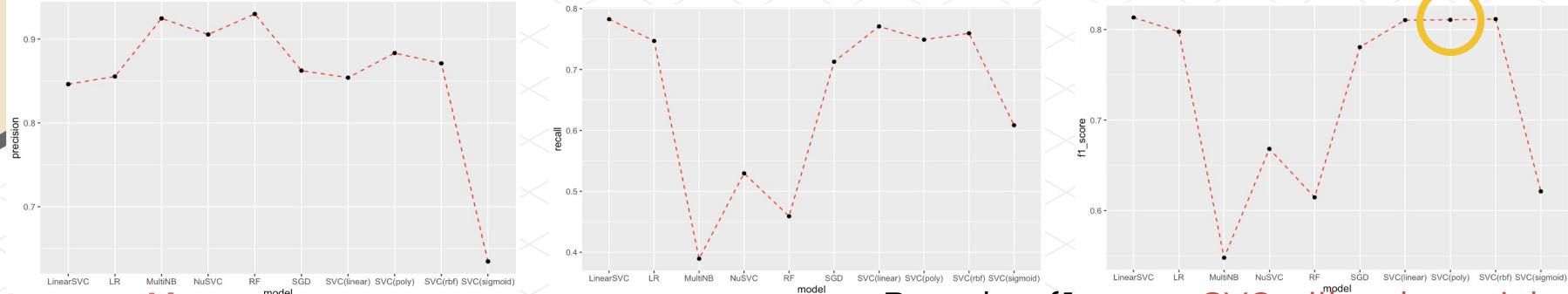
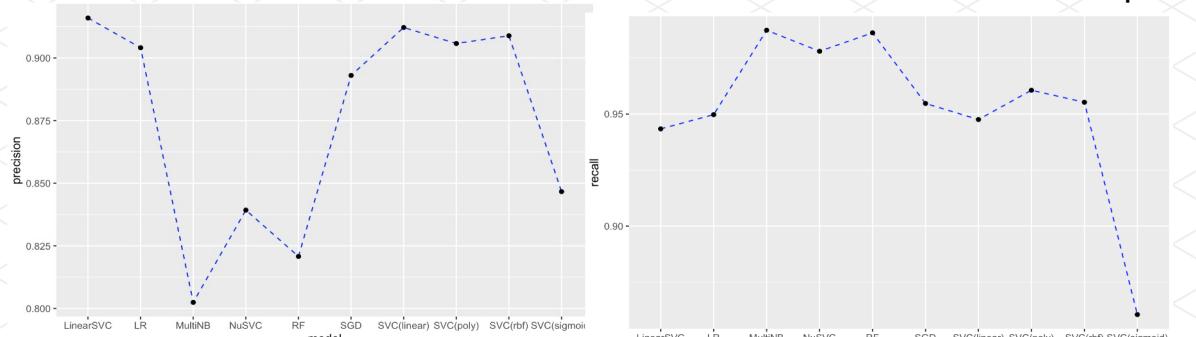
Negative



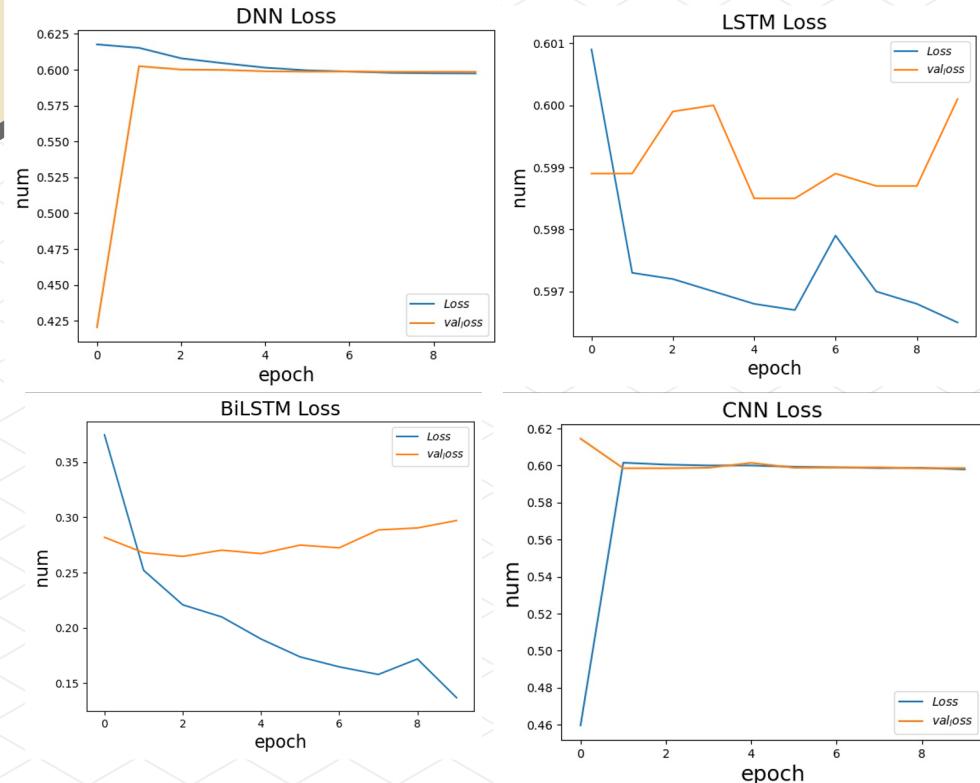
Positive



Based on f1-score, SVC with polynomial kernel performs the best on validation set

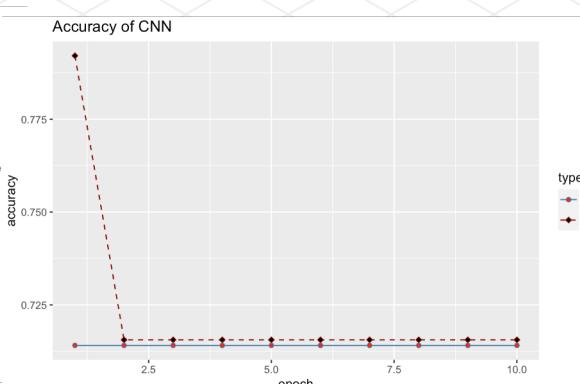
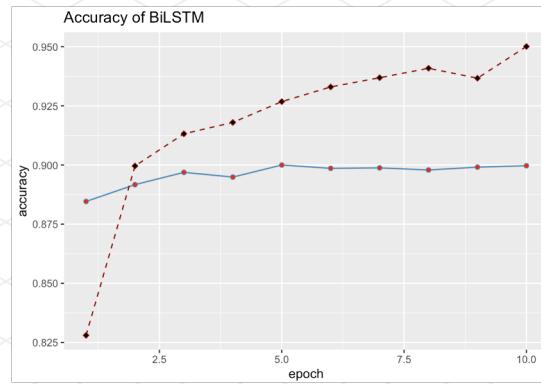
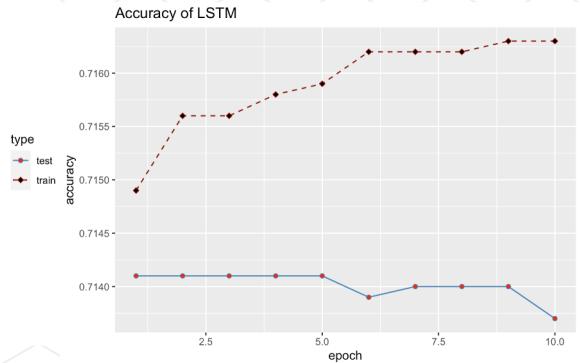
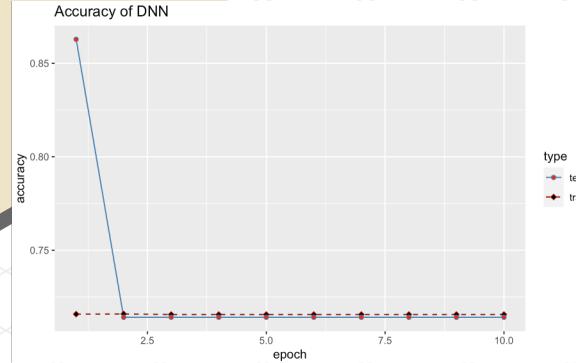


# Model evaluation-Deep Learning Models



Based on loss, **BiLSTM** performs the best on validation set among these four deep learning models

# Model evaluation-Deep Learning Models



BiLSTM performs the best on validation set among these four deep learning models with 89.97% accuracy.

However, as SVC with polynomial kernel has higher accuracy, we choose to use SVC with polynomial kernel as our final model

# Parameter tuning

SVC (polynomial kernel) with C  
= 1.0 and degree = 2

Accuracy: 90.04%



Accuracy: 90.48%



# 05

# Outcomes

# Flow Chart of Demo

## Input

Product asin



## Data Crawling

Crawl 1000 reviews of this product



## Data Pre-processing

Perform the same with the process as when training.



# Flow Chart of Demo

## Predict by model

Predict the sensitive attributes of the reviews



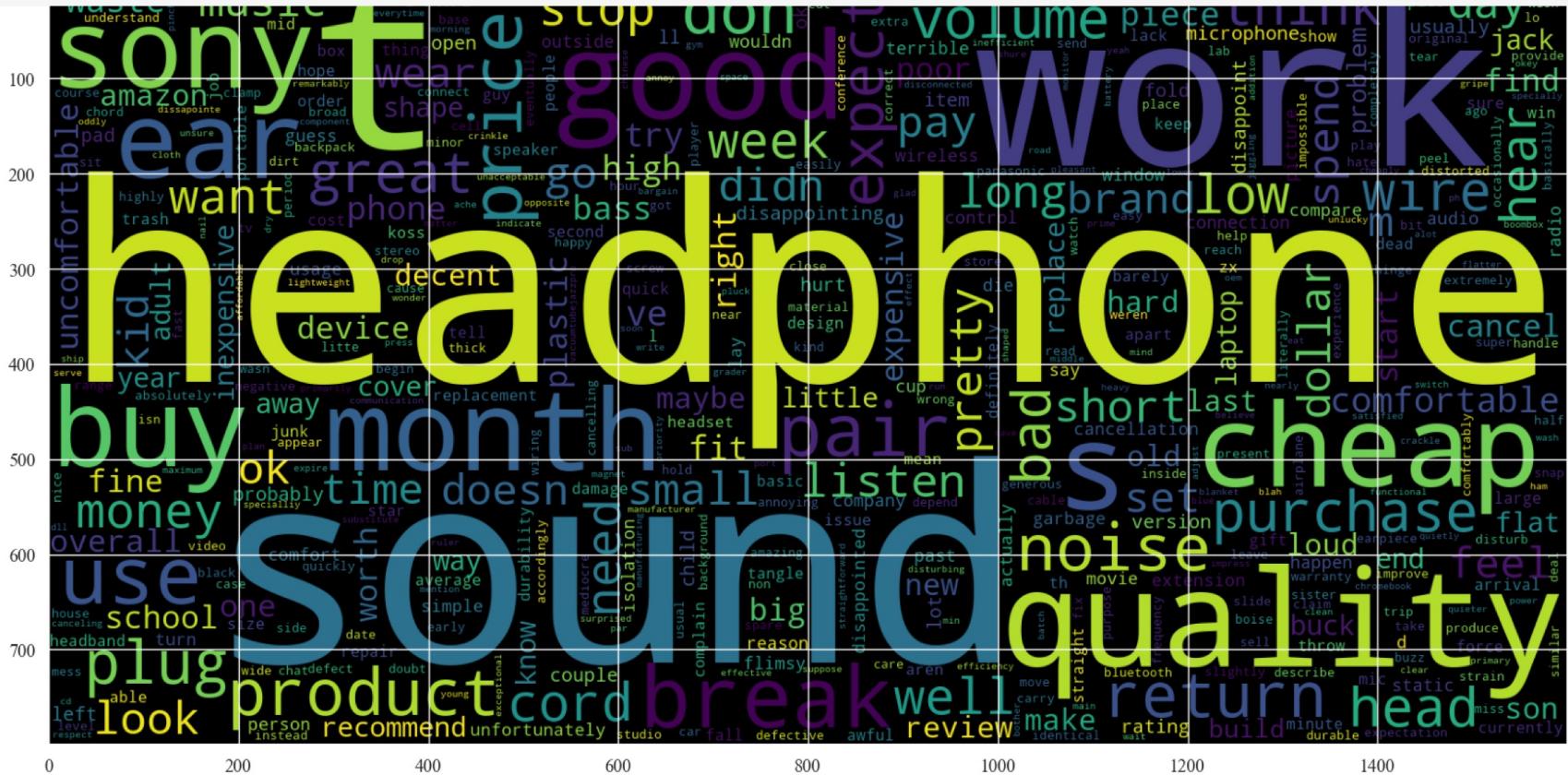
## Output

Visualize in table, figure and Word Cloud.



[Let's see the demo!](#)

# Demo



# Demo Output

```
▶ 
product_asin = "B00NJ2M43M"
a = amazon_senti()
a.fair_senti_table(product_asin)
```

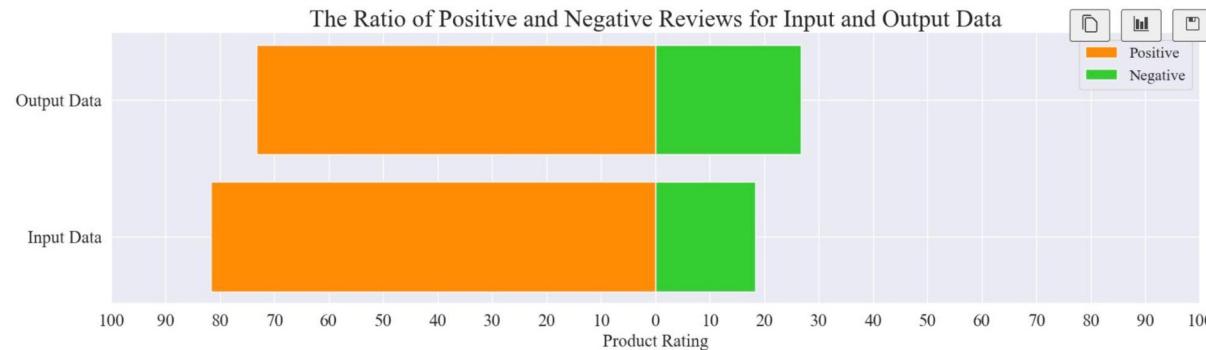
[147] ✓ 0.5s

```
...
      Input    Output
-----
Positive 81.7% 73.3%
Negative 18.3% 26.7%
```

```
▶ 
a.fair_senti_visual(product_asin)
```

7] ✓ 0.1s

Python



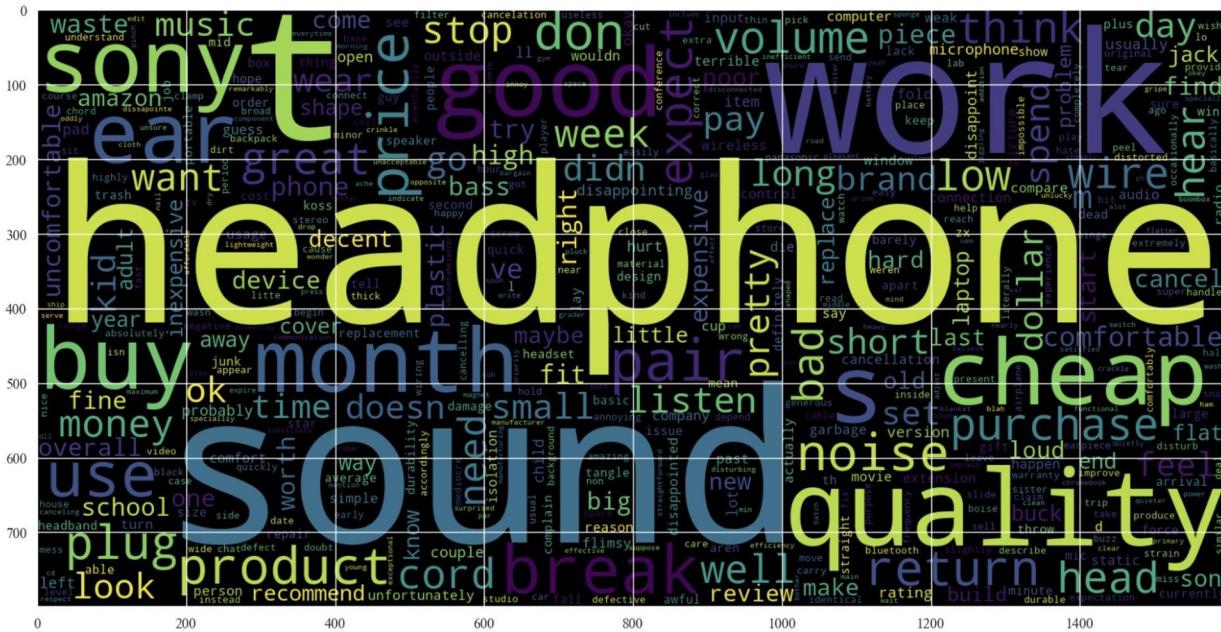
# Demo Output

```
a.fair_senti_wordcloud(product_asin)
```

✓ 2.9s

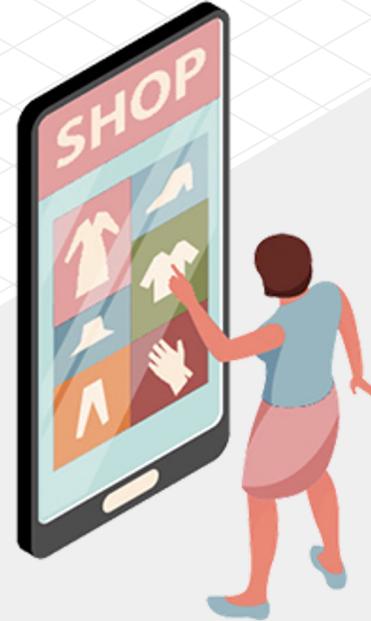
Python Python

<matplotlib.image.AxesImage at 0x2b4f032b0>



# 06

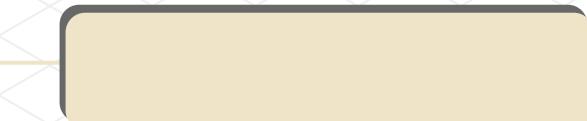
# Conclusion



# Summarize

Data Pre-  
Processing

Demo



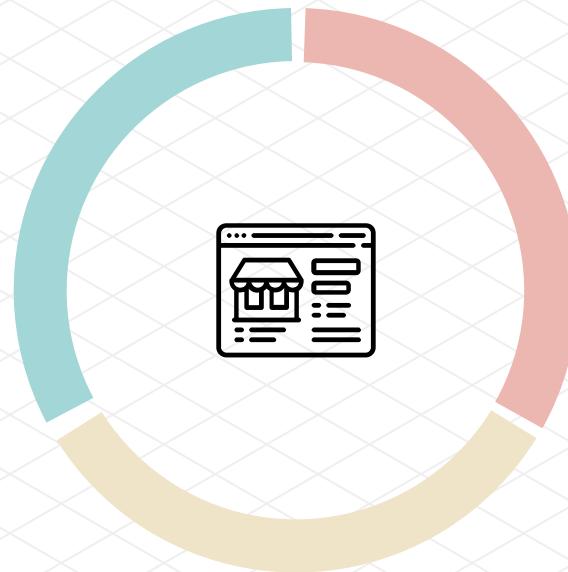
Review Crawling

Model Selecting

# Conclusion

## Summarize

1. Review Crawling
2. Data cleaning
3. Match the best model
4. Demo



## Compared with the proposal

1. Complete all the ideas mentioned.
2. Optimize the presentation form of the Output.

## Future Prospect

Can be applied to all Amazon products!!!



## Limitation

- The first situation is that we cannot completely avoid merchants manually reviewing positive and negative reviews, because generally positive reviews and negative reviews are machine reviews, which can have a certain similarity in text information. However, if the merchant asks users who have purchased this product to give a good review, we cannot predict this situation.
- Secondly, we only did parameter tuning for the model with the best performance in validation, but not for other models. After adjusting the parameters, other models may have better performance.

## References:

- [1] Rodrigues, A. P., Fernandes, R., A, A., B, A., Shetty, A., K, A., Lakshmanna, K., & Shafi, R. M. (2022). Real-time Twitter spam detection and sentiment analysis using machine learning and Deep Learning Techniques. *Computational Intelligence and Neuroscience*, 2022, 1–14. <https://doi.org/10.1155/2022/5211949>
- [2] Nagi Alsubari, S., N. Deshmukh, S., Abdullah Alqarni, A., Alsharif, N., H. H. Aldhyani, T., Waselallah Alsaade, F., & I. Khalaf, O. (2021). Data Analytics for the identification of fake reviews using supervised learning. *Computers, Materials & Continua*, 70(2), 3189–3204. <https://doi.org/10.32604/cmc.2022.019625>
- [3] Fang, X., Zhan, J. Sentiment analysis using product review data. *Journal of Big Data* 2, 5 (2015). <https://doi.org/10.1186/s40537-015-0015-2>
- [4] Nguyen, Heidi; Veluchamy, Aravind; Diop, Mamadou; and Iqbal, Rashed (2018) "Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches," SMU Data Science Review: Vol. 1: No. 4, Article 7. <https://scholar.smu.edu/datasciencereview/vol1/iss4/7>

# THANKS

Do you have any questions?

