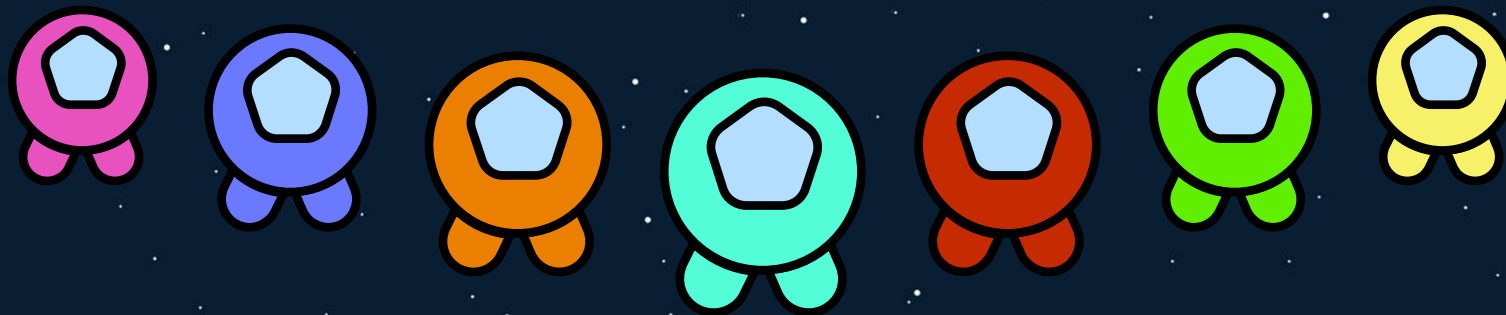


NLP ON AMAZON REVIEWS

There are ? **negative reviews** among us

Group 02 Lecture A



Mengchen Qu, Yiheng Zhang, Siyao Lu.

TABLE OF CONTENTS

01 Data Description

02 Data Preprocessing
Topic Models(LDA)

03 Sentiment Analysis
(Multi-Class)

04 Sentiment Analysis
(Binary)

12 -> 5

variables

ID, RATING, REVIEW, SUMMARY

100 -> 0

NAs in "Review" Variable

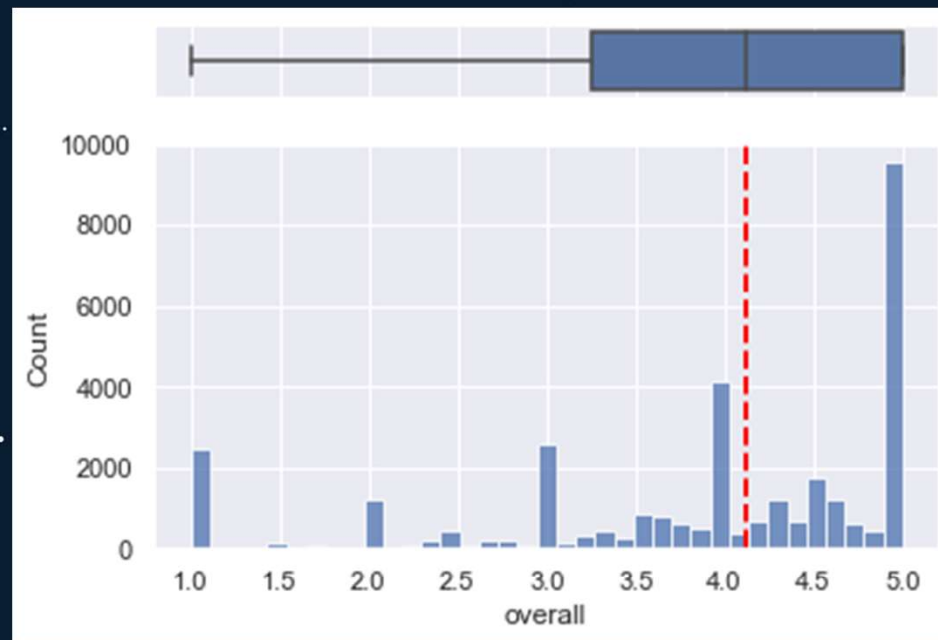
2,565,349 -> 500,000

rows

DATA VISUALIZATION

Rating

Review Length



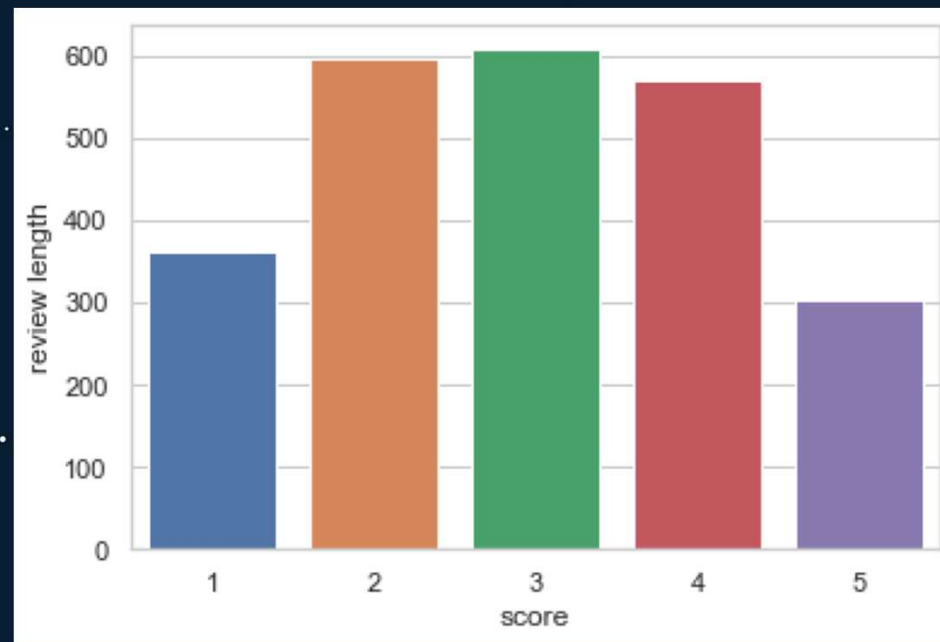
Review
(wordcloud)

Review
(bar chart)

DATA VISUALIZATION

Rating

Review Length



Review
(wordcloud)

Review
(bar chart)

[illegible]

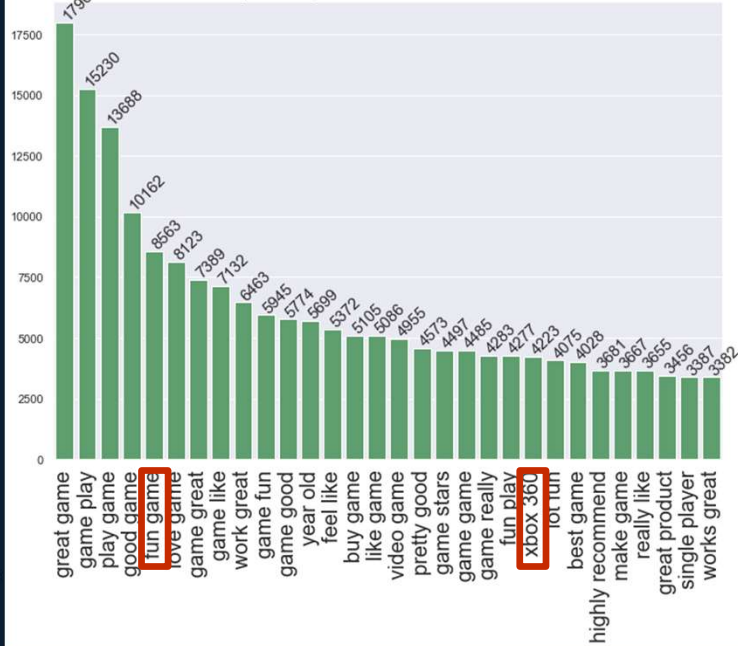
Review (bar chart)



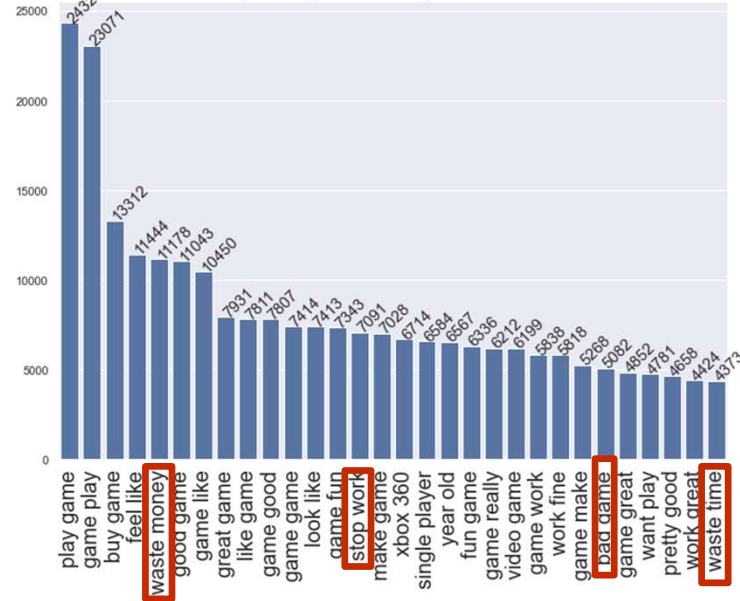
Review (bar chart)

DATA VISUALIZATION

Top 20 Bigrams in Positive Reviews



Top 20 Bigrams in Negative Reviews



Review
(wordcloud)

Review
(bar chart)

PREPROCESSING

I'm a big fan of the Brainwavz S1 (actually all of their headphones have ye
ith any of their products). The S1 has been my main set for active use (e.g., workouts, r

× special characters

× accents

stopwords removal



PREPROCESSING

× special characters



× accents



stopwords removal



lemmatization

Lower & tokenization

1. Review -> sentences
2. Remove punctuations
3. Sentences -> words
4. Nouns, verbs, adjectives, adverbs -> root
5. list -> string

PREPROCESSING

(naïve naive) or (fiancé fiancée)

× special characters

× accents

stopwords removal



PREPROCESSING

× special characters



× accents



stopwords removal



lemmatization



Lower & tokenization



PREPROCESSING

- x special characters



x accents



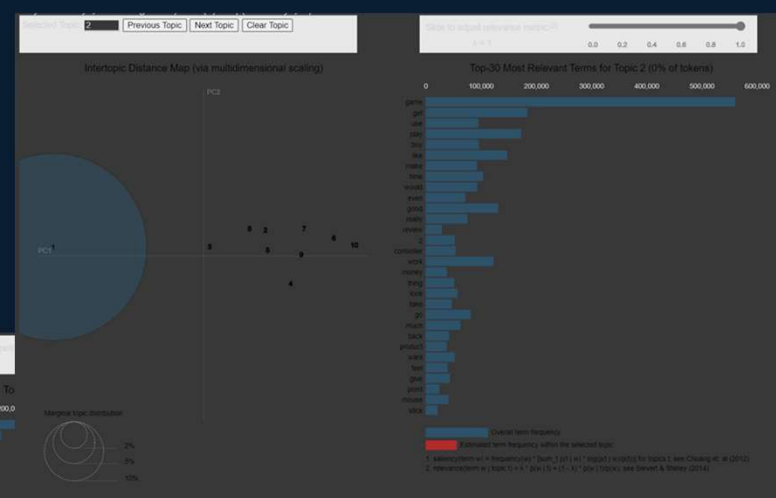
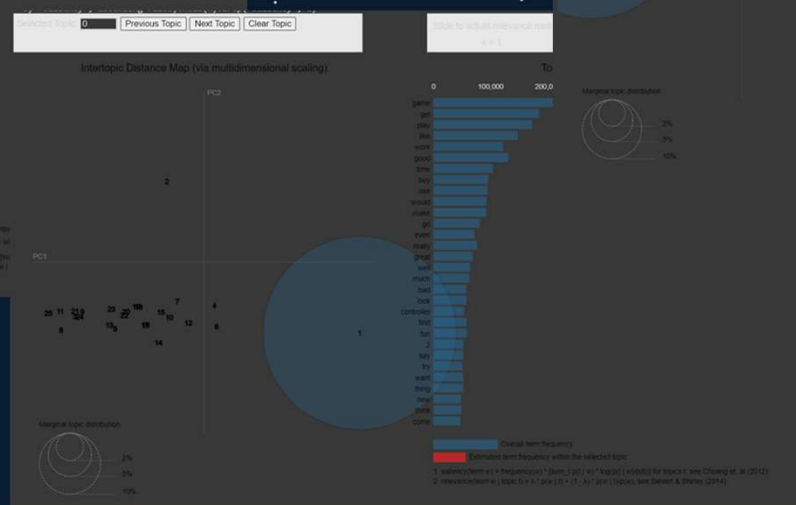
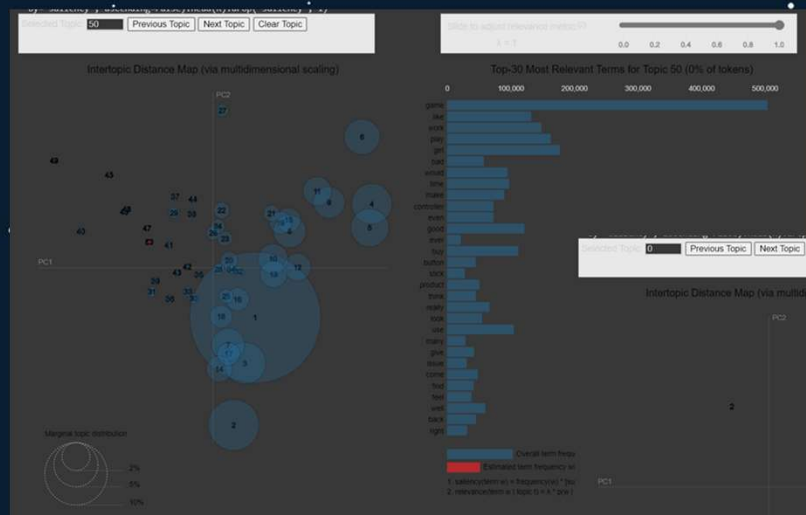
stopwords removal



```
stopwords = nltk.corpus.stopwords.words('english')
stopwords += list(string.punctuation)
stopwords += ["'", '"', '...', '`', ',', '(', ')', ':', ';', '!', '@',
'said', "'s", "also", 'one', "n't", 'com', 'satirewire', '-', '-',
'-', '_', 'satirewire.com', '/"]
```

TOPIC MODEL — — — RATING ≤ 3

Unigram + num_topics = 50/ 25/ 10



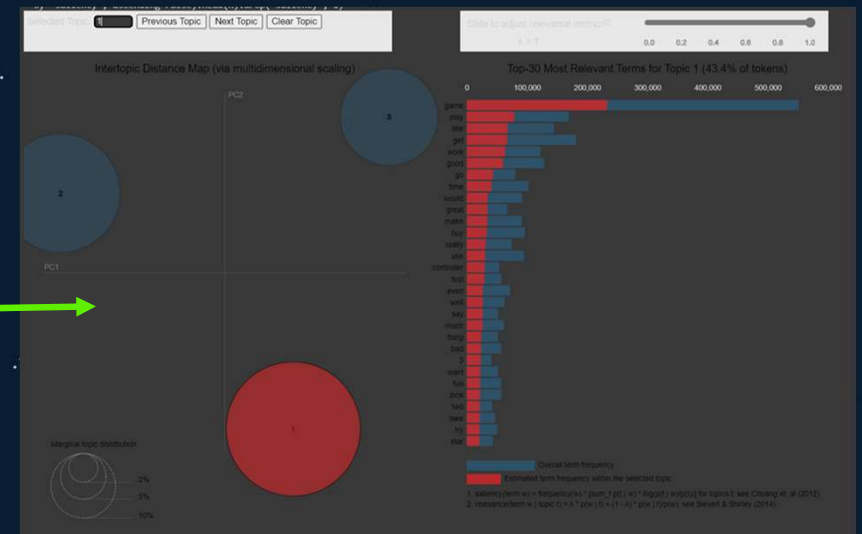
Unigram/ Bigram/ Trigram + num_topics = 3

Topics:

1. Controller Problem
2. Mouse and other accessories problems
3. Bad Games

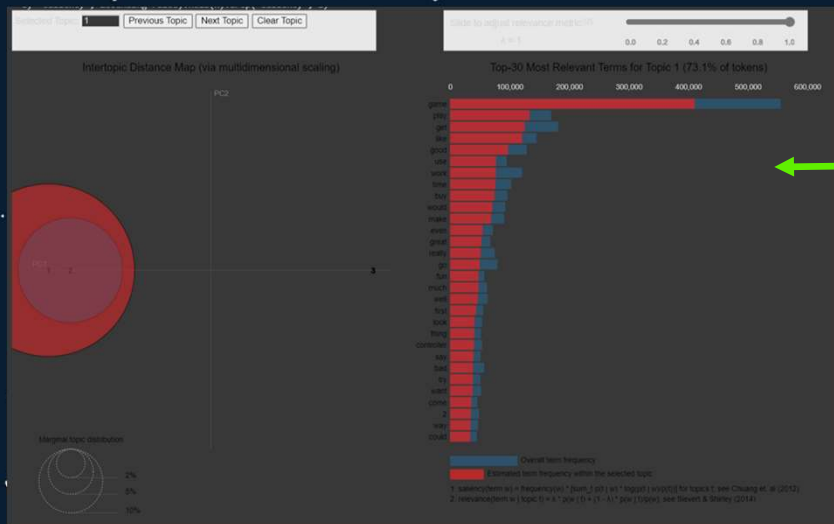


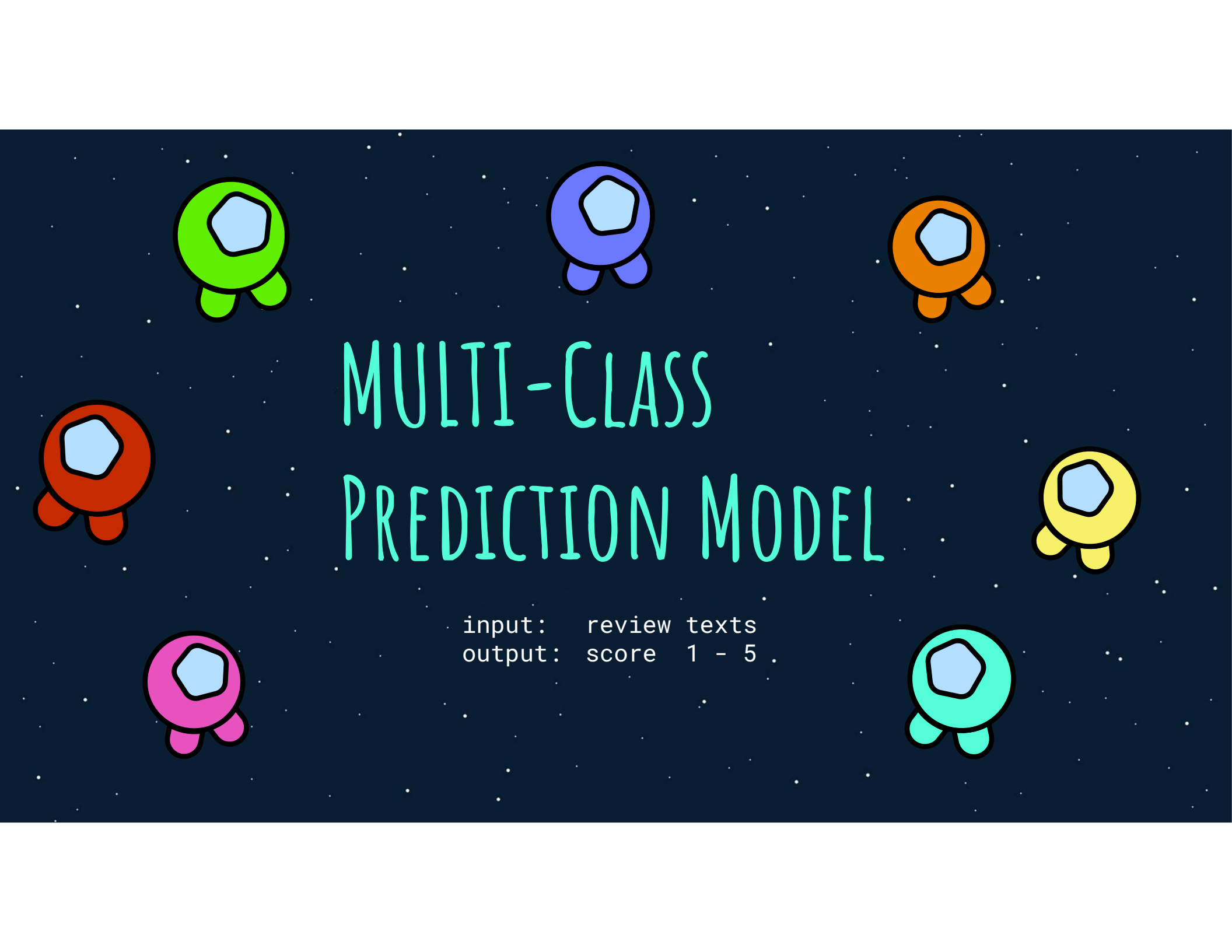
Trigram



Bigram

Unigram



The background is a dark blue space filled with small white stars. Seven colorful cartoon rockets are scattered around the central text. The rockets are green, purple, orange, red, yellow, pink, and cyan. Each rocket has a light blue hexagonal window and two small legs.

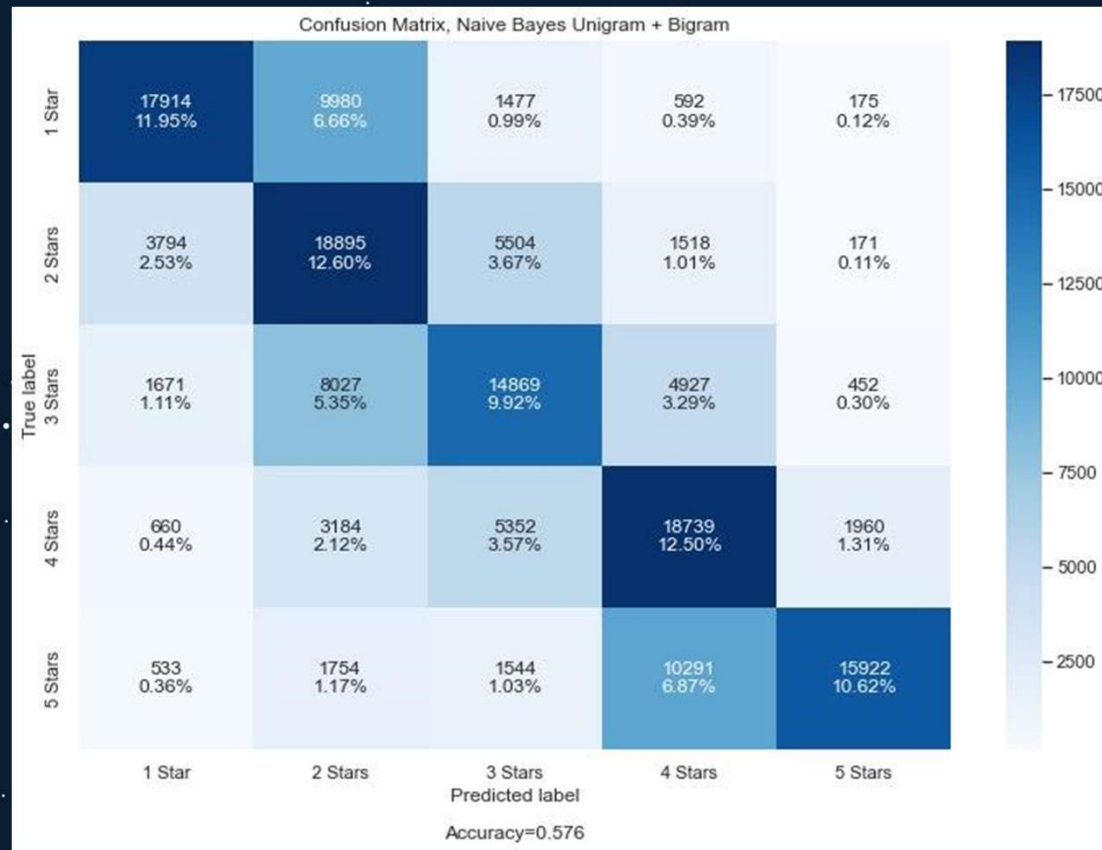
MULTI-CLASS PREDICTION MODEL

input: review texts
output: score 1 - 5

MODEL WITH BALANCED DATA

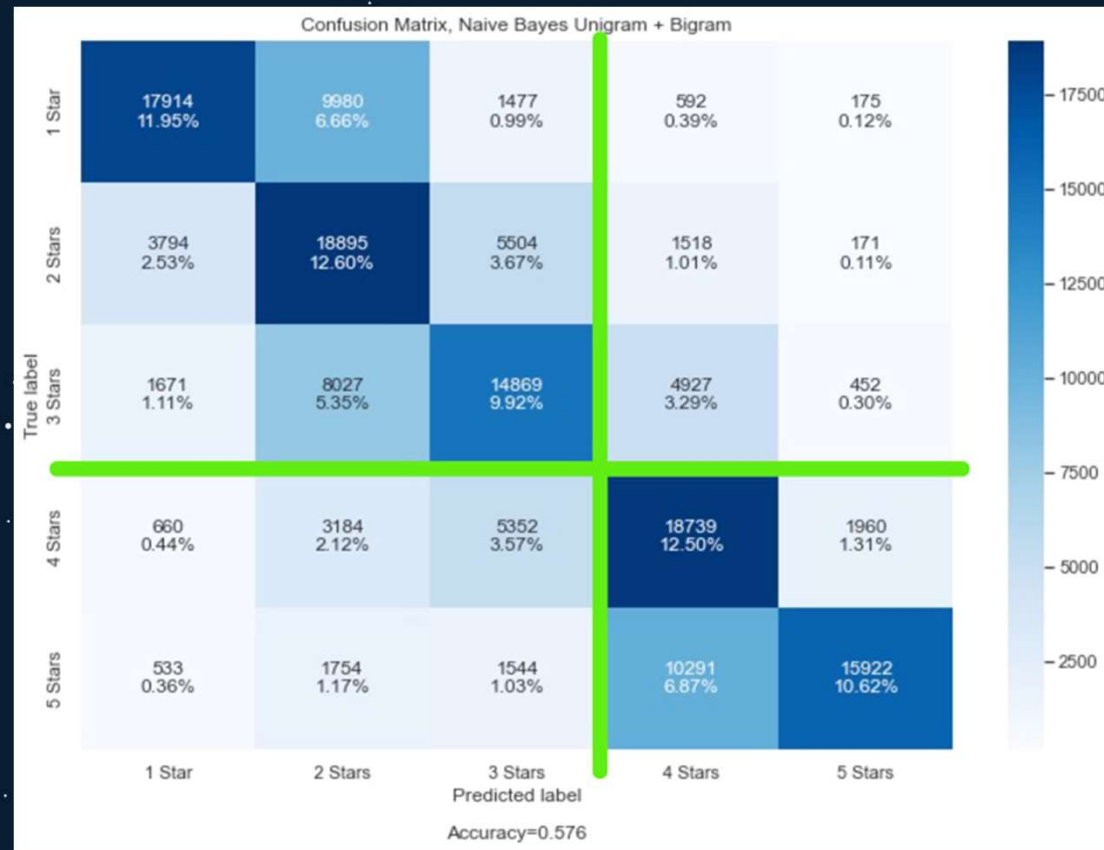
	TFIDF + Unigram	TFIDF + Uni +Bigram
Naive Bayes	0.559	<u>0.576</u>
Linear Regression	0.436	N/A

★ BEST MULTICLASS MODEL TEST RESULTS ★

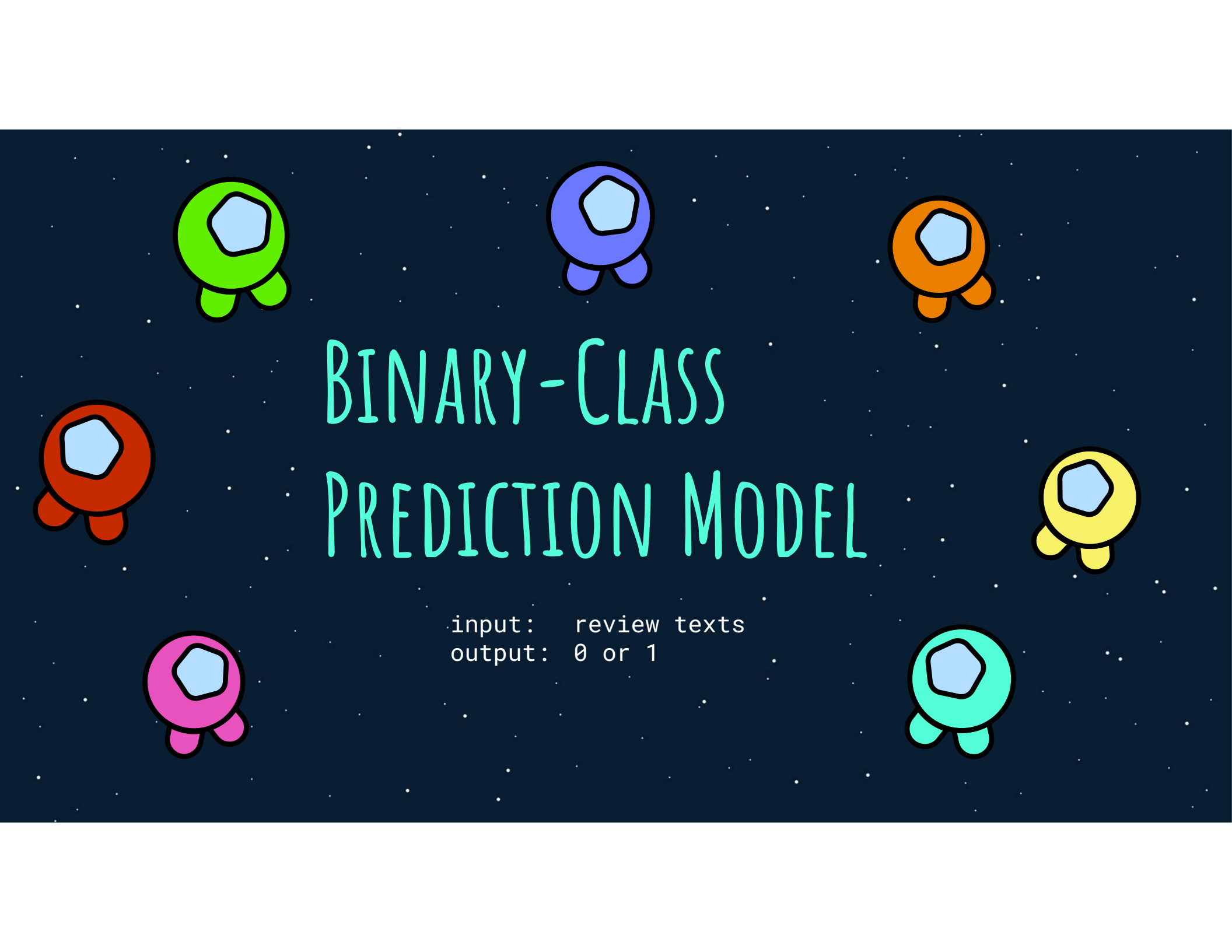


TFIDF Naive Bayes Unigram + Bigram

≤ 3 stars



≥ 4 stars



BINARY-CLASS PREDICTION MODEL

input: review texts
output: 0 or 1

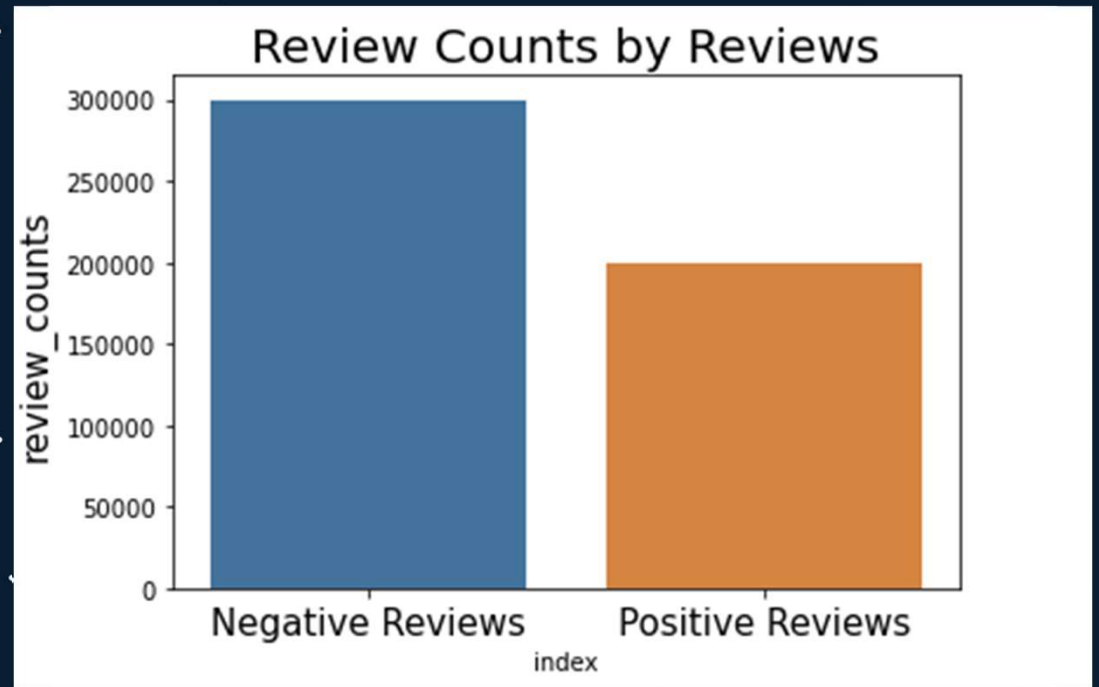
WHAT DID WE DO?



Binary Classification

Positive: Rating ≥ 4

Negative: Rating < 4



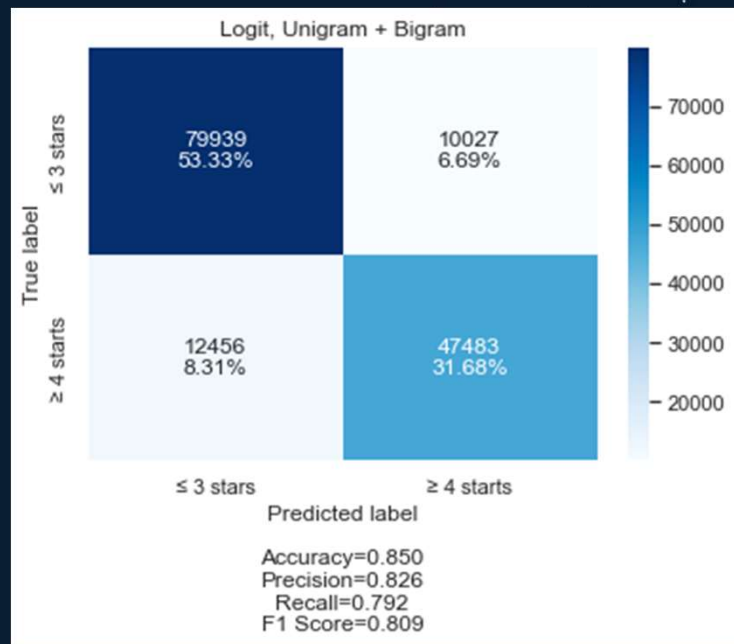
MODEL WITHOUT OVERSAMPLE

	TFIDF + Unigram	TFIDF + Uni +Bigram	BOW + Unigram	BOW + Uni +Bigram
Naive Bayes	0.783	0.736	0.804	0.821
Logistic Regression	0.839	<u>0.850</u>	0.828	0.842

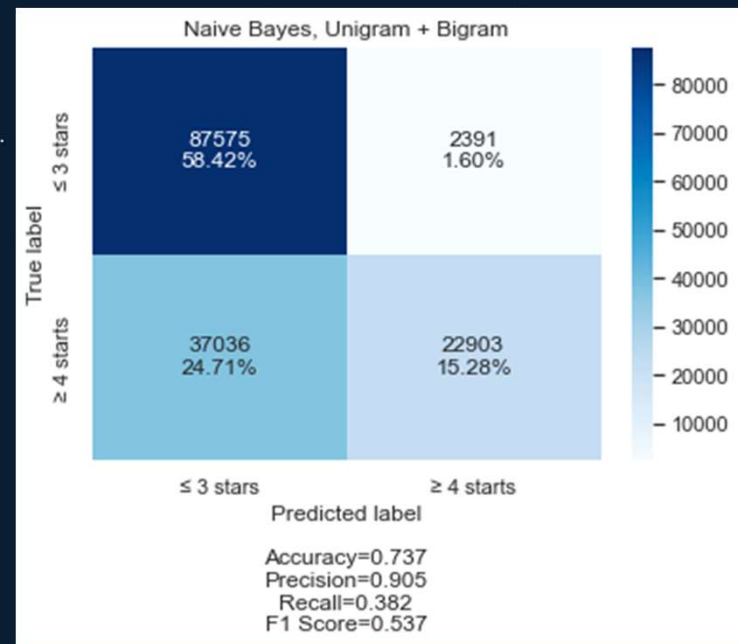
MODEL WITH OVERSAMPLE

	TFIDF + Unigram	TFIDF + Uni +Bigram	BOW + Unigram	BOW + Uni +Bigram
Naive Bayes	0.803	0.808	0.789	0.819
Logistic Regression	0.826	<u>0.837</u>	0.814	0.831

BEST BINARY MODEL TEST RESULTS



TFIDF Logistic
Unigram + Bigram





TFIDF Naive Bayes
Unigram + Bigram







amazon prime Deliver to Yiheng Irvine 92612 All Hello, Yiheng Account & Lists Returns & Orders Cart

All Customer Service Groceries Amazon Basics Home Improvement Pet Supplies Beauty & Personal Care Handmade Health & Household Coupons


Create Review

 Logitech G603 LIGHTSPEED Wireless Gaming Mouse, HERO 12K Sensor, 12,000 DPI, Lightweigh...


Overall rating



Rate features
For gaming  
Ergonomic  
Battery life  

Add a headline

Add a photo or video
Shoppers find images and videos more helpful than text alone.


Add a written review

Choose your public name
This is how you'll appear to other customers
 Yiheng Zhang
Don't worry, you can always change this on your profile



overall rating

The target the model predicts;

headline & written review

Contents for NLP;

The inputs the model takes in;..

Create Review



Logitech G603 LIGHTSPEED Wireless Gaming Mouse, HERO 12K Sensor, 12,000 DPI, Lightweigh...

Overall rating



Submitted Clear

Add a headline

Great shape, no latency, lightweight (104g), and as affordable as all other gaming mouse

Add a photo or video

Shoppers find images and videos more helpful than text alone.



Add a written review

Pretty much perfect. With one Energizer Ultimate Lithium battery, the weight is just 104g which is light enough. I recommend buying your own batteries unless you like heavy mice. On high mode, there is no latency at all. Coming from a Deathadder, the shape of the G603 is better and gives you a bit more space to be accurate. I have medium-large hands and it fits just about perfectly. It's large enough to be comfortable and not have my hands slide along the mousepad, but also just small enough to be able to make the small micro-adjustments with my fingers to be accurate. I instantly lowered my sensitivity in Overwatch from 5 in-game to 4.5 and I was more accurate than before.

Case 1: Negative rating by mistake

- The customer
 - gives 2 stars
 - starts writing reviews
- The prediction model
 - starts analyzing
 - predicts sentiment continuously
 - identifies positive sentiment
 - reports confliction
 - negative rating
 - positive review

Create Review



Logitech G603 LIGHTSPEED Wireless Gaming Mouse, HERO 12K Sensor, 12,000 DPI, Lightweigh...

Overall rating



Is this the rating you want to give?

Submitted Clear

Add a headline

Great shape, no latency, lightweight (104g), and as affordable as all other gaming mouse

Add a photo or video

Shoppers find images and videos more helpful than text alone.



Add a written review

Pretty much perfect. With one Energizer Ultimate Lithium battery, the weight is just 104g which is light enough. I recommend buying your own batteries unless you like heavy mice. On high mode, there is no latency at all. Coming from a Deathadder, the shape of the G603 is better and gives you a bit more space to be accurate. I have medium-large hands and it fits just about perfectly. It's large enough to be comfortable and not have my hands slide along the mousepad, but also just small enough to be able to make the small micro-adjustments with my fingers to be accurate. I instantly lowered my sensitivity in Overwatch from 5 in-game to 4.5 and I was more accurate than before.

Case 1: Negative rating by mistake

- The customer
 - gives 2 stars
 - starts writing reviews
- The prediction model
 - starts analyzing
 - predicts sentiment continuously
 - identifies positive sentiment
 - reports confliction
 - negative rating
 - positive review

Solution: a pop-up reminder

Create Review



Logitech G603 LIGHTSPEED Wireless Gaming Mouse, HERO 12K Sensor, 12,000 DPI, Lightweigh...

Overall rating



Submitted Clear

Add a headline

Super common scrolling issues

Add a photo or video

Shoppers find images and videos more helpful than text alone.



Add a written review

Worked great for a day then the scroll wheel stopped working properly. It scrolls in the opposite direction 1/3 the time randomly, making the scroll wheel useless. I downloaded updated firmware made specifically to fix "intermittent scrolling issues". Didn't help at all and I noticed that there were TONS of others with the same problem. Save yourself the trouble. Buy something that works.



Case 2: Common issue

- The LDA model continuously
 - analyze reviews of the product
 - summarize common issues
- The customer
 - gives 2 stars
 - starts writing reviews
- The LDA model
 - starts analyzing
 - detect common issue "scroll"

Create Review



Logitech G603 LIGHTSPEED Wireless Gaming Mouse, HERO 12K Sensor, 12,000 DPI, Lightweigh...

Overall rating



Submitted Clear

Add a headline

Super common scrolling issues

Add a photo or video

Shoppers find images and videos more helpful than text alone.



See [Logitech mouse scrolling issue](#)

Add a written review

Worked great for a day then the scroll wheel stopped working properly. It scrolls in the opposite direction 1/3 the time randomly, making the scroll wheel useless. I downloaded updated firmware made specifically to fix "intermittent scrolling issues". Didn't help at all and I noticed that there were TONS of others with the same problem. Save yourself the trouble. Buy something that works.



Case 2: Common issue

- The LDA model continuously
 - analyze reviews of the product
 - summarize common issues
- The customer
 - gives 2 stars
 - starts writing reviews
- The LDA model
 - starts analyzing
 - detect common issue "scroll"

Solution:

- pop-up links to solution
- report to manufacturer

SUMMARY

01

Data Preprocessing

02

LDA Analysis

03

Sentiment Analysis

04

Application



THANK YOU

BANA 275 Group 02

Siyao Lu
Mengchen Qu
Yiheng Zhang