

The background of the slide is a complex network graph. It consists of numerous circular nodes of varying sizes, connected by thin lines. The nodes are colored in dark grey, red, and teal. The lines connecting them are primarily dark grey, with some red lines highlighting specific clusters or paths. The overall layout is split vertically: the left half has a dark grey background, and the right half has a light blue background.

# Predicción de evaluaciones a partir de reseñas

**Proyecto Final**

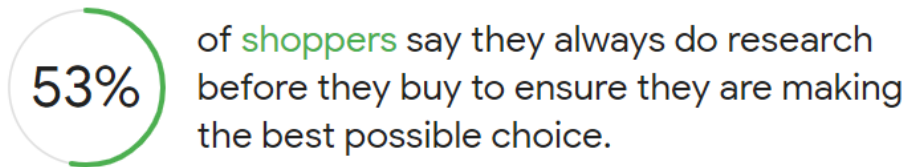
**Por: Kevin Knights**

**Dataset: Amazon Customer Reviews**

---

# Identificación de Problema

- La mayoría consumidores investigan y consultan distintos medios de información antes de realizar una compra.
- Aproximadamente 2 de 3 consumidores utilizan videos en línea para realizar su compra.



Nearly 2/3 of shoppers say online video has given them ideas and inspiration for their purchase.

# Identificación de Problema

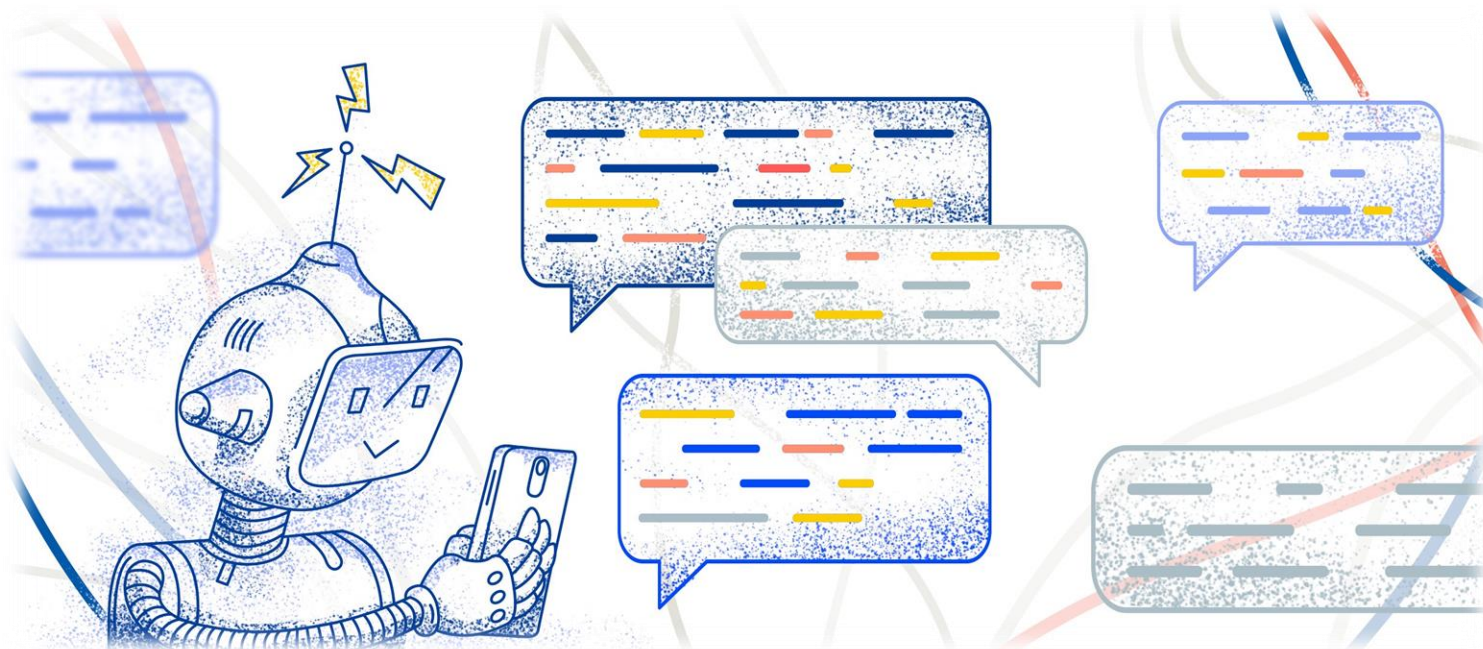
¿Cómo los vendedores pueden cuantificar toda esta información en forma de comentarios y artículos para entender la satisfacción del cliente?

Respuesta: NLP (Natural Language Processing)

---



# NLP en una imagen



NLP es una rama de la Inteligencia Artificial, que permite a las computadoras entender el lenguaje humano a partir de texto, y con esto realizar distintas tareas.

# Dataset

Para entrenar nuestro modelo utilizamos un conjunto de datos de reseñas de Amazon, especialmente de la categoría de electrónicos.

	star_rating	review_headline	review_body
0	5.0	Battery life is excellent. Pairs with my phone...	Battery life is excellent. Pairs with my phon...
1	4.0	Good, but.	Good sound quality, but occasional fuzzy inter...
2	1.0	These over priced, poor quality headphones don...	My wife bought me these headphones on Amazon o...
3	5.0	pretty great.	It's an HDMI cable. It conducts HDMI signals f...
4	4.0	Four Stars	its somewhat ok not easy to use

# Limpieza y Transformación

- Debido a que nuestro data es texto, debemos realizar los siguientes pasos para limpiarla.
    1. Convertir el texto a minúscula (unificación).
    2. Remover signos de puntuación.
    3. Remover caracteres especiales (incluyendo emojis).
    4. Remover posibles números y urls.
    5. Remover palabras relleno.
-

# Limpieza y Transformación

## Antes de la limpieza

	review	star_rating
0	Battery life is excellent. Pairs with my phone...	5.0
1	Good, but. Good sound quality, but occasional ...	4.0
2	These over priced, poor quality headphones don...	1.0
3	pretty great. It's an HDMI cable. It conducts ...	5.0
4	Four Stars its somewhat ok not easy to use	4.0

## Después de la limpieza

	review	star_rating
0	battery life excellent pairs phone tablet easi...	5.0
1	good good sound quality occasional fuzzy inter...	4.0
2	priced poor quality headphones dont even deser...	1.0
3	pretty great hdmi cable conducts hdmi signals ...	5.0
4	four stars somewhat ok easy use	4.0

# Limpieza y Transformación

- Limpiar el texto no es suficiente, debemos reducir las palabras a su mínima expresión usando una de las siguientes técnicas:

## Stemming vs Lemmatization

change  
changing  
changes  
changed  
changer

chang

change  
changing  
changes  
changed  
changer

change



# Limpieza y Transformación

## Antes del Stemming

	review	star_rating
0	battery life excellent pairs phone tablet easi...	5.0
1	good good sound quality occasional fuzzy inter...	4.0
2	priced poor quality headphones dont even deser...	1.0
3	pretty great hdmi cable conducts hdmi signals ...	5.0
4	four stars somewhat ok easy use	4.0

## Después del Stemming

	review	star_rating
0	batteri life excel pair phone tablet easili ba...	5.0
1	good good sound qualiti occasion fuzzi interfe...	4.0
2	price poor qualiti headphon dont even deserv s...	1.0
3	pretti great hdmi cabl conduct hdmi signal dev...	5.0
4	four star somewhat ok easi use	4.0

# Limpieza y Transformación

- Aún la data no es útil para entrenar un modelo, porque es texto.
- Para que sea entendible por la computadora debemos convertirlo a una matriz.

Document Term Matrix

	intelligent	applications	creates	business	processes	bots	are	i	do	intelligence
Doc 1	2	1	1	1	1	0	0	0	0	0
Doc 2	1	1	0	0	0	1	1	0	0	0
Doc 3	0	0	0	1	0	0	0	1	1	1

# Limpieza y Transformación

En este caso se utilizo una técnica llamada TF-IDF para generar la matriz.  
Donde TF-IDF significa: Term Frequency – Inverse Data Frequency

$$w_{i,j} = t f_{i,j} \times \log \frac{N}{d f_j}$$

# occurrences of term in document

# total documents

tf-idf score

# documents containing word

Term document matrix with tf-idf			
words\documents	Document1	document2	query term
cat	0	0	0
runs	0	0	0
behind	0	0	0
rat	0.30103	0	0.30103
dog	0	0.30103	0

# Limpieza y Transformación

DTM (Document Term Matrix) resultante truncada a las 15 mil columnas mas relevantes y la variable objetivo “star\_rating”.

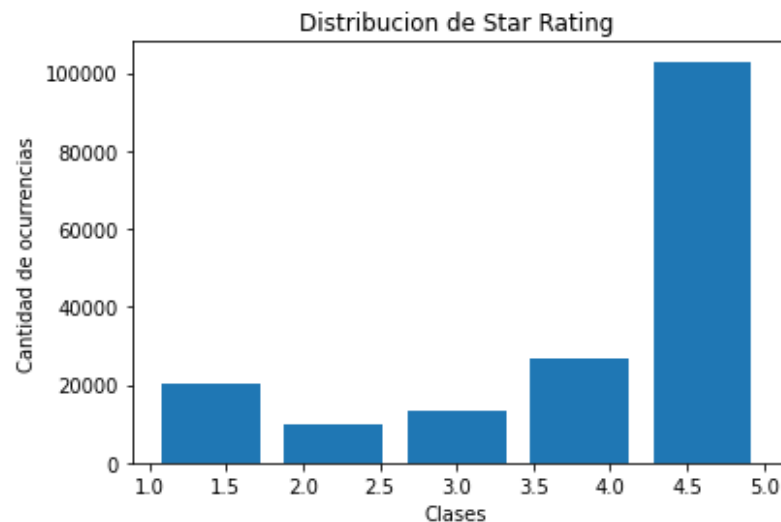
```
dtm_with_target.head()
```

	star	five	great	work	good	sound	product	one	price	use	...	reunion	boseawesom	laptopcomput	een	performavc	control	sooooooooo	desnt	phil	star_rating
0	0.000000	0.0	0.078611	0.000000	0.000000	0.000000	0.109684	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0
1	0.000000	0.0	0.000000	0.000000	0.344172	0.170667	0.000000	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0
2	0.095948	0.0	0.049750	0.096098	0.000000	0.000000	0.069415	0.0	0.138611	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
3	0.000000	0.0	0.192372	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0
4	0.201862	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.252538	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0

5 rows × 15001 columns

# Análisis Exploratorio

- Como la mayoría de problemas de clasificación nuestra data esta desbalanceada.

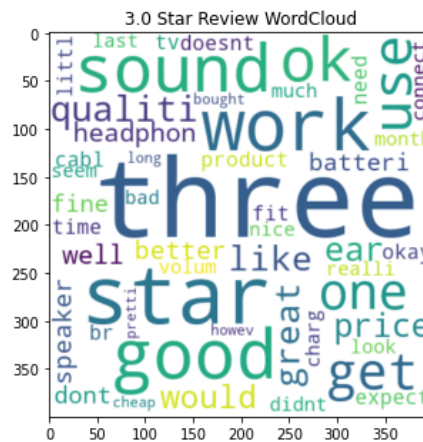
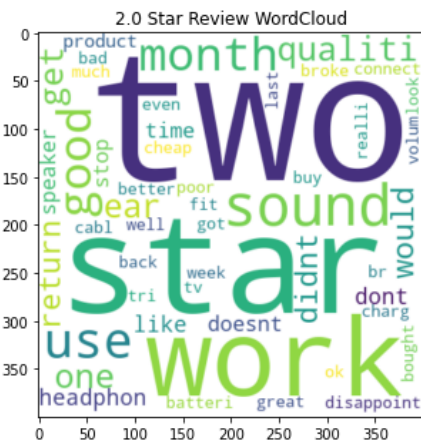
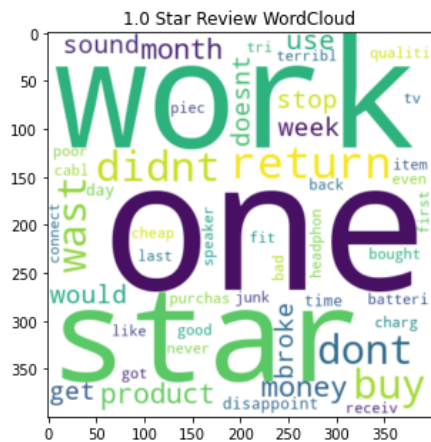


['1 : 20293', '2 : 9733', '3 : 13153', '4 : 26924', '5 : 102927']

star_rating	
count	173030.000000
mean	4.054493
std	1.395862
min	1.000000
25%	4.000000
50%	5.000000
75%	5.000000
max	5.000000

# Análisis Exploratorio

- ¿Qué palabras son las más relevantes por clase (reseña)?





# Análisis Exploratorio

- ¿Cuál es el vocabulario (palabras únicas) de cada clase (reseña)?

```
1.0 star review Vocabulary | # Items: 9 | Items: ['money', 'never', 'piec', 'day', 'receiv', 'junk', 'first', 'wast', 'terribl']
2.0 star review Vocabulary | # Items: 1 | Items: ['two']
3.0 star review Vocabulary | # Items: 4 | Items: ['howev', 'long', 'three', 'okay']
4.0 star review Vocabulary | # Items: 9 | Items: ['devic', 'small', 'far', 'job', 'bluetooth', 'four', 'set', 'problem', 'bit']
5.0 star review Vocabulary | # Items: 14 | Items: ['exactli', 'recommend', 'ship', 'perfect', 'perfectli', 'happi', 'awesom', 'five', 'best',
'fast', 'amaz', 'excel', 'thank', 'instal']
```



# Entrenamiento

- El modelo seleccionado para esta clasificación fue el Linear Support Vector Classification (Linear SVC).
  - Este modelo es bastante común dentro de este tipo de aplicaciones.
  - Este modelo se encuentra dentro de la librería de scikit-learn.
-

# Evaluación - Métricas

- Validación cruzada con K-Fold (k-fold = 5):

`Cross_validation Metrics - LinearSVC: [0.77457107 0.78183132 0.77590753 0.77905003 0.77976448]`

`Cross_validation Mean - LinearSVC: 0.7782248854215995`

- Precisión (Accuracy):

`Accuracy: 0.7771484713633474`

- Cohen Kappa:

`Cohen Kappa Metric: 0.5972339037906819`

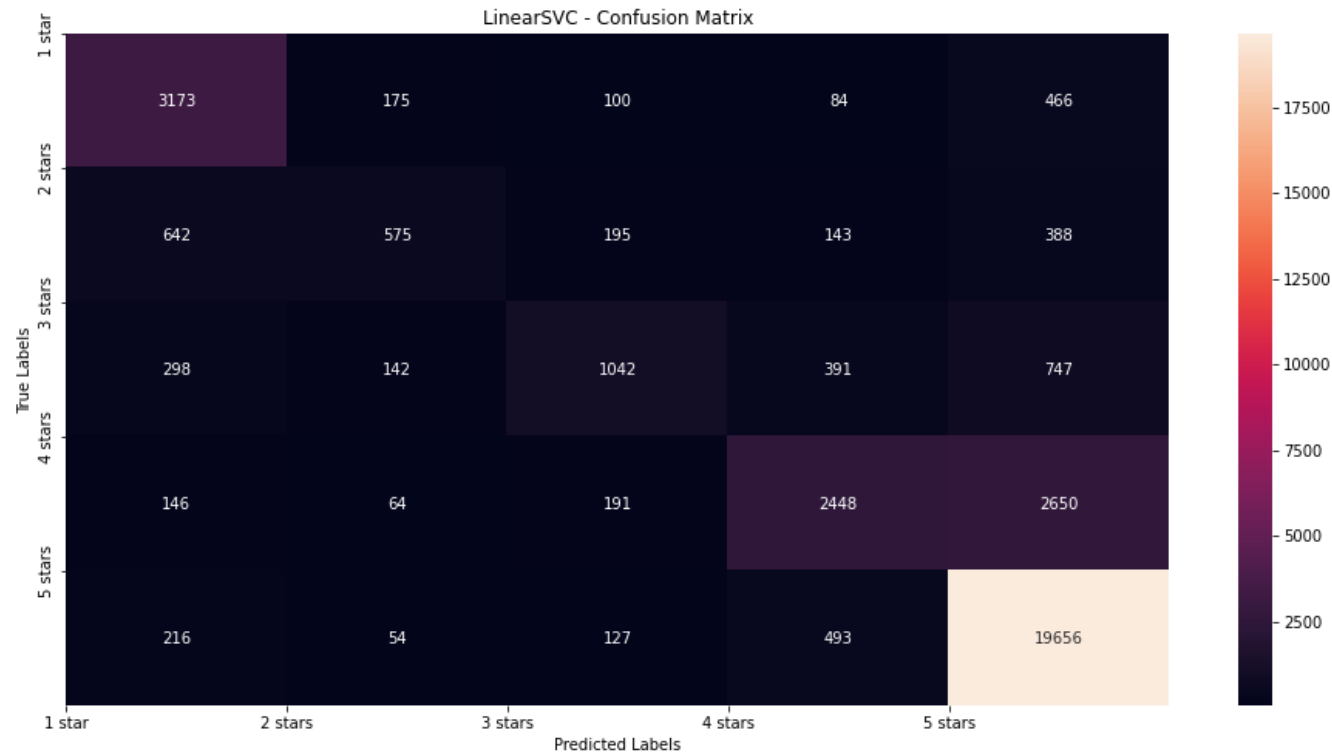
- Matthews Correlation Coefficient:

`Matthews Correlation Coefficient MCC Metric: 0.6079991160689276`

---

# Evaluación - Métricas

- Matriz de confusión:



# Evaluación - Métricas

- Reporte de Clasificación:

```
print(classification_report(y_test, y_pred, target_names=labels))
```

	precision	recall	f1-score	support
1 : 20293	0.71	0.79	0.75	3998
2 : 9733	0.57	0.30	0.39	1943
3 : 13153	0.63	0.40	0.49	2620
4 : 26924	0.69	0.45	0.54	5499
5 : 102927	0.82	0.96	0.88	20546
accuracy			0.78	34606
macro avg	0.68	0.58	0.61	34606
weighted avg	0.76	0.78	0.76	34606

# Predicciones

- Seleccionamos un producto cualquiera dentro de la categoría de electrónicos y consultamos distintas fuentes de información :

[Back to results](#)



Roll over image to zoom in

Razer DeathAdder Essential Gaming Mouse: 6400 DPI Optical Sensor - 5 Programmable Buttons - Mechanical Switches - Rubber Side Grips - Classic Black

Visit the Razer Store  
★★★★★ 10,407 ratings  
42 answered questions

Amazon's Choice for "razer ..."

List Price: ~~\$49.99~~ Details  
Price: **\$19.99** + \$21.02 shipping  
You Save: **\$30.00 (60%)**  
\$26.08 Shipping & Import Fees Deposit to Panama [Details](#)

0:02 / 0:35

The Razer DeathAdder Essential

413,761 views · Jul 9, 2018

6.3K DISLIKE SHARE SAVE ...

**RAZER**  
1.96M subscribers

Meet the new Razer DeathAdder Essential (<http://rzt.to/dae-e>), sporting our multi-award winning Razer Mechanical Mouse Switches, co-developed with Omron, a precise 6,400 DPI optical sensor and iconic ergonomics that are tried and tested for more than a decade.

[SHOW MORE](#)

**BESTMOUSE.COM** BEST MICE BEST BUDGET MOUSE REVIEWS BLOG Search the site

## Razer DeathAdder Essential Gaming Mouse Review

The Razer DeathAdder Essential **gaming mouse** is a much more price effective access to the full-sized DeathAdder collection. It has the exact same shape and size as the much more price versions, and it's suitable for all hold kinds for almost all hand dimensions, although people with little hands could have problem with a claw or fingertip grip.

It has a reduced minimum lift-off range, excellent click latency, and a relatively consistent sensor, but the CPI range isn't very wide, and also you can change the CPI in increments of 100 just. Sadly, the computer mouse is fairly heavy, its braided wire isn't extremely versatile, and also its backlighting is environment-friendly just.

Page Contents



# Predicciones



Scott Ryan

★☆☆☆☆ **Junk, worst scroll wheel I've ever used on a mouse**

Reviewed in the United States on November 25, 2021

Color: Classic Black | Style: New Model | **Verified Purchase**

I've had this thing for like a month and the scroll wheel is BUSTED, every time I try to scroll down a page the scroll wheel decides it has a mind of it's own and just scrolls up and down however which way it feels. What prompted me to write this review was about 10 minutes ago when I was attempted to scroll down a website and the mouse literally just took me to the top of the page and then continued to scroll up over and over while I was scrolling down on the scroll wheel. Spend a little more money and get something you won't have to replace almost immediately, this product is junk.

Helpful

Report abuse

```
review_title = "Junk, worst scroll wheel I've ever used on a mouse"
review_body = "I've had this thing for like a month and the scroll wheel is BUSTED, every time I try to scroll down a page the scroll wheel de

review = review_title + ' ' + review_body

predictReviewStars(review)
```

Input Review: Junk, worst scroll wheel I've ever used on a mouse I've had this thing for like a month and the scroll wheel is BUSTED, every t  
ime I try to scroll down a page the scroll wheel decides it has a mind of it's own and just scrolls up and down however which way it feels. W  
hat prompted me to write this review was about 10 minutes ago when I was attempted to scroll down a website and the mouse literally just took  
me to the top of the page and then continued to scroll up over and over while I was scrolling down on the scroll wheel. Spend a little more m  
oney and get something you won't have to replace almost immediately, this product is junk.  
Predicted Stars: [1.]

# Predicciones



John S.

★★★★☆ Like the tracking for gaming. Build quality meh?

Reviewed in the United States on November 28, 2021

Color: Classic Black | Style: Original Model | **Verified Purchase**

I guess for tracking this mouse works better than some regular desktop mice I have owned. Not impressed having a light up logo on body but some may find that attractive. Didn't like the software recommended to download. I decided not to install it for the few games I played I felt I did not need to customizing. I like the mouse size and weight and the fabric cord was nice touch. Build quality didn't seem like anything special. Some say they started getting double clicks, but I thought the clicks were fine?

Helpful

| Report abuse


```
review_title = "Like the tracking for gaming. Build quality meh?"
review_body = "I guess for tracking this mouse works better than some regular desktop mice I have owned. Not impressed having a light up logo
review = review_title + ' ' + review_body
```

```
predictReviewStars(review)
```

```
Input Review: Like the tracking for gaming. Build quality meh? I guess for tracking this mouse works better than some regular desktop mice I
have owned. Not impressed having a light up logo on body but some may find that attractive. Didn't like the software recommended to download.
I decided not to install it for the few games I played I felt I did not need to customizing. I like the mouse size and weight and the fabric
cord was nice touch. Build quality didn't seem like anything special. Some say they started getting double clicks, but I thought the clicks w
ere fine?
```

```
Predicted Stars: [3.]
```

# Predicciones

[BEST MICE](#)[BEST BUDGET MOUSE](#)[REVIEWS](#)[BLOG](#)

## Razer DeathAdder Essential Gaming Mouse Review

The Razer DeathAdder Essential gaming mouse is a much more price effective access to the full-sized DeathAdder collection. It has the exact same shape and size as the much more price versions, and it's suitable for all hold kinds for almost all hand dimensions, although people with little hands could have problem with a claw or fingertip grip.

It has a reduced minimum lift-off range, excellent click latency, and a relatively consistent sensor, but the CPI range isn't very wide, and also you can change the CPI in increments of 100 just. Sadly, the computer mouse is fairly heavy, its braided wire isn't extremely versatile, and also its backlighting is environment-friendly just.

Page Contents

### RAZER DEATHADDER ESSENTIAL REVIEW

Input Review: Razer DeathAdder Essential Gaming Mouse Review

The Razer DeathAdder Essential gaming mouse is a much more price effective access to the full-sized DeathAdder collection. It has the exact same shape and size as the much more price versions, and it's suitable for all hold kinds for almost all hand dimensions, although people with little hands could have problem with a claw or fingertip grip.

It has a reduced minimum lift-off range, excellent click latency, and a relatively consistent sensor, but the CPI range isn't very wide, and also you can change the CPI in increments of 100 just. Sadly, the computer mouse is fairly heavy, its braided wire isn't extremely versatile, and also its backlighting is environment-friendly just.

The Razer DeathAdder Important is decent for office usage. It really feels sturdy and also is suitable for mostly all hand dimensions with any type of hold kind, but little hands might have problem with a claw or fingertip hold. You can program all of its buttons making use of the Razer Synapse 3 software, which is readily available on Windows just. Regrettably, its scroll wheel does not have any kind of L/R tilt switches, as well as it can not open for free scrolling.

The Razer DeathAdder Essential benefits FPS gaming. It has a low minimum CPI, a reduced lift-off distance, and also great click latency, although it might not feel responsive enough for some gamers. It feels well-built and appropriates for almost all hand sizes using any grip kind, yet people with small hands may have problem with a claw or fingertip grip. Regrettably, it's rather heavy, and also its cable television isn't very versatile. Also, you can change the CPI in increments of 100 only.

The Razer DeathAdder Necessary is a great MMO gaming mouse, although it doesn't have almost as lots of switches as dedicated MMO mice. Nonetheless, you can program all of them, as well as you can likewise set a HyperShift switch to make it possible for a 2nd layer of commands. It has a low minimum lift-off distance as well as excellent click latency, although it may not really feel responsive sufficient for some gamers.

It feels durable as well as comfortable, and also it appropriates for mostly all hand sizes utilizing any type of hold type, although little hands may struggle with a claw or fingertip hold. Sadly, its cord isn't really adaptable, as well as you can adjust the CPI in increments of 100 only.

The Razer DeathAdder Necessary is a good ultra-light gaming mouse, although it's much larger than what we typically consider an ultra-light mouse. Nevertheless, this sturdy-feeling version appropriates for the majority of hand sizes with any hold type, however people with small hands may fight with a claw or fingertip grasp. It has PTFE feet which provide a good gliding experience as well as great click latency, although it might not really feel responsive sufficient to some players. Sadly, its cable television isn't very flexible.

The Razer DeathAdder Necessary is inadequate for traveling given that it's bulkier than most mice developed for traveling and also likely will not fit in many laptop cases. Likewise, you can only use it with its cable, which may be troublesome in limited spaces like on public transport.

The ballot rate options on the Razer DeathAdder Necessary are 500Hz as well as 1000Hz. The sensor on our unit is fairly constant, but it often tends to overshoot the set CPI a lot more when moving the computer mouse gradually than when release quick.

Predicted Stars: [5.]

# Predicciones



**Dinky** 3 years ago

I'm not satisfied with my Razer DeathAdder Elite both rubber grips have fallen off, and the back mouse feet started to peel off.



REPLY

```
review_title = "The Razer DeathAdder Essential"
review_body = "I'm not satisfied with my Razer DeathAdder Elite both rubber grips have fallen off, and the back mouse feet started to peel off."
review = review_title + ' ' + review_body

predictReviewStars(review)
```

Input Review: The Razer DeathAdder Essential I'm not satisfied with my Razer DeathAdder Elite both rubber grips have fallen off, and the back mouse feet started to peel off.  
Predicted Stars: [5.]

# Conclusiones

- El modelo aun puede ser optimizado mejorando sus hyperparametros.
  - Balanceando el dataset podemos esperar mejores resultados.
  - El dataset de entrenamiento fue delimitado por motivos de recursos computacionales, podemos estimar que con un mayor numero de registro y columnas del DTM el modelo tendrá un mayor poder predictivo.
  - Considero que el modelo obtuvo un buen rendimiento, tomando en cuenta las limitantes anteriores.
-