



Text analytics

Miguel Acosta
Angela Jimenez
Andrés Rincón

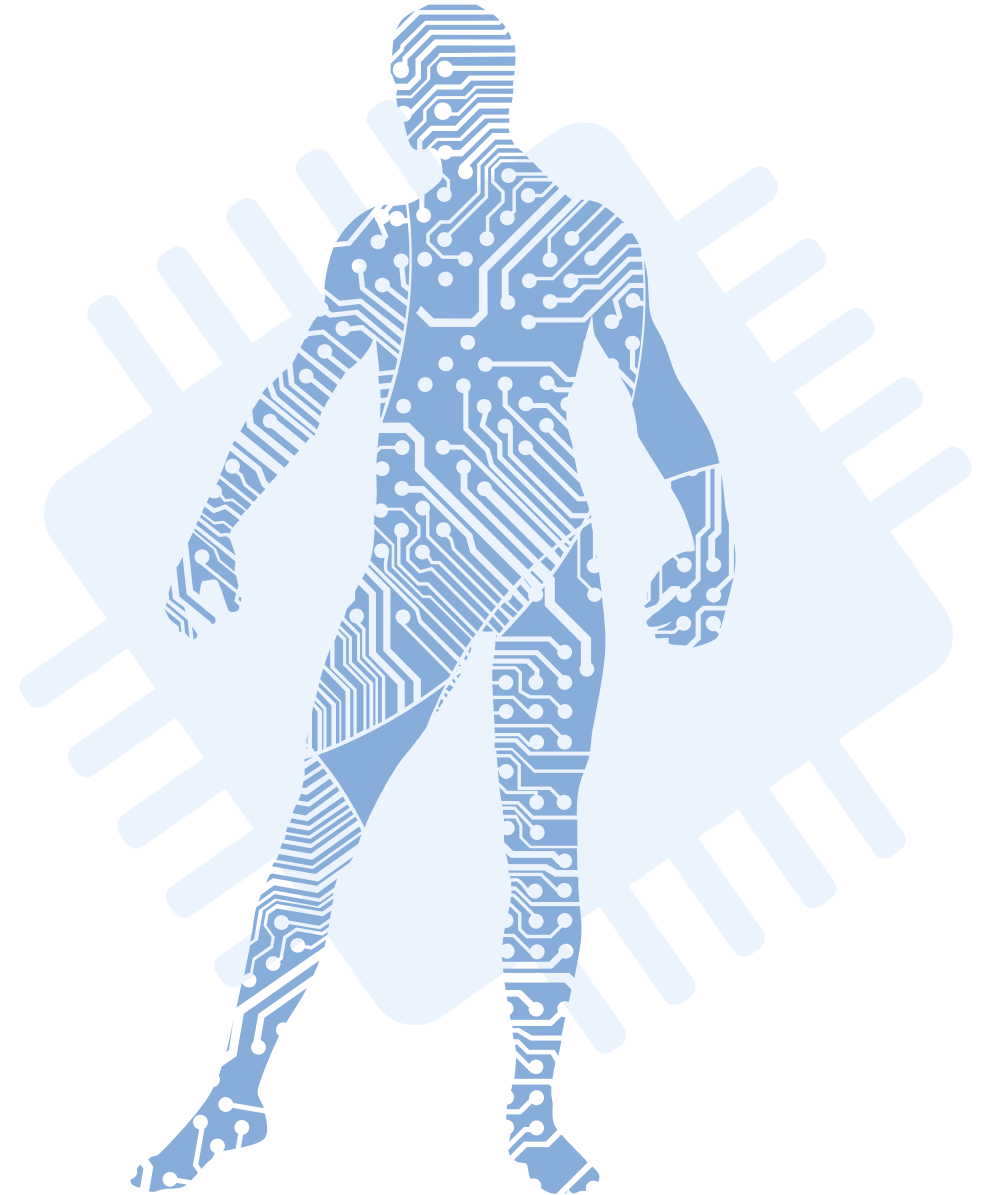
Agenda

01 Análisis de datos

02 Modelos de ML

03 Tableros de control

04 Conclusiones



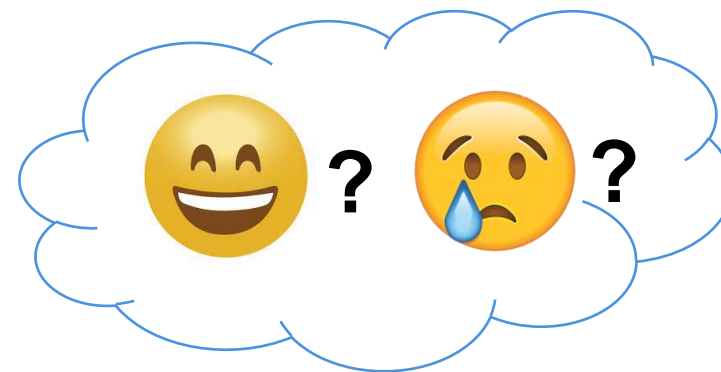
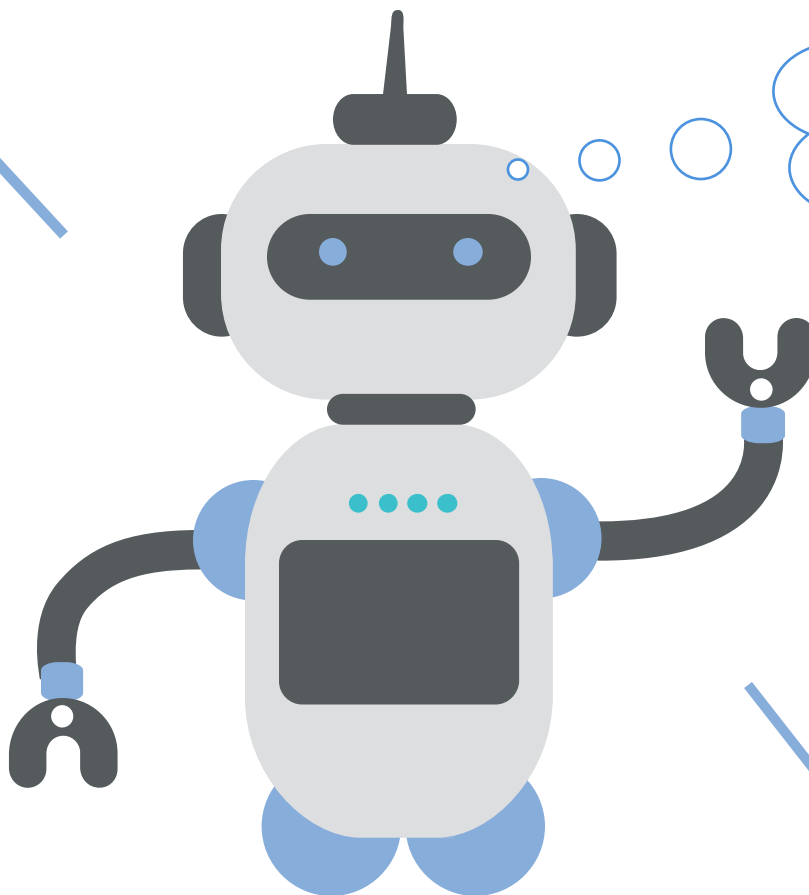


Objetivos



Crear un clasificador

Queremos crear un modelo que sea capaz de clasificar las emociones de un post

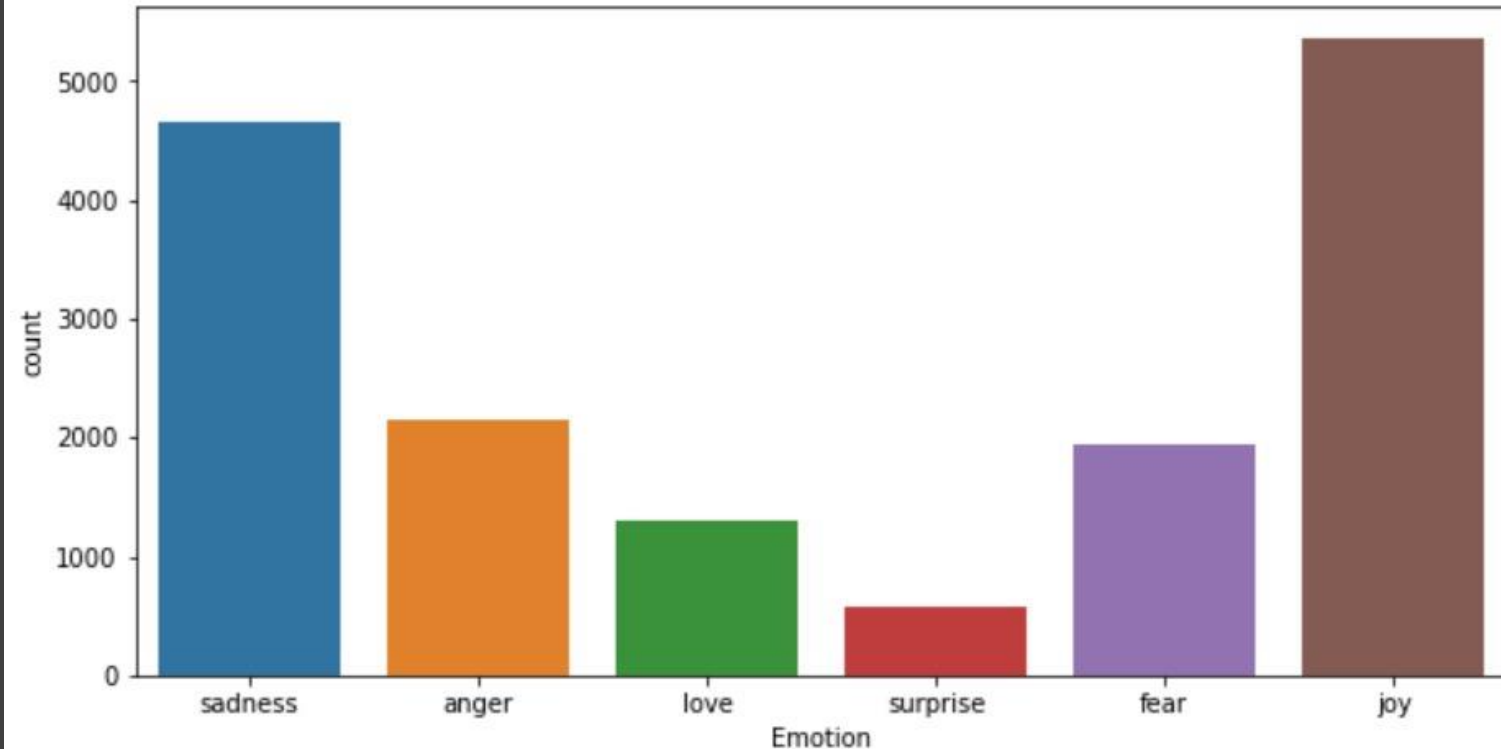


Opciones de mejora del negocio

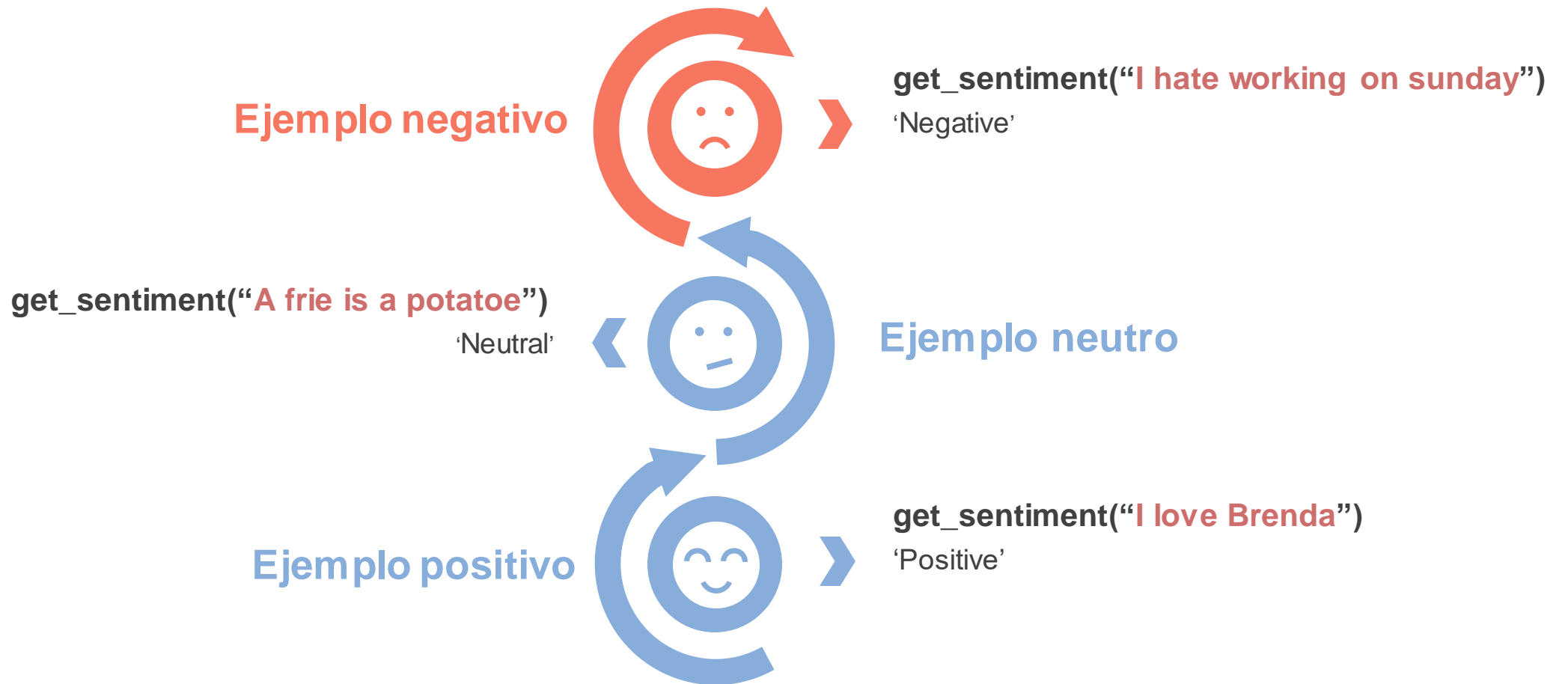
Usar los resultados obtenidos por el perfilamiento de datos y el modelo obtenido para aportar al negocio.

Análisis de datos de entrada

Emoción	Conteo
sadness	4666
anger	2159
love	1304
surprise	572
fear	1937
joy	5362

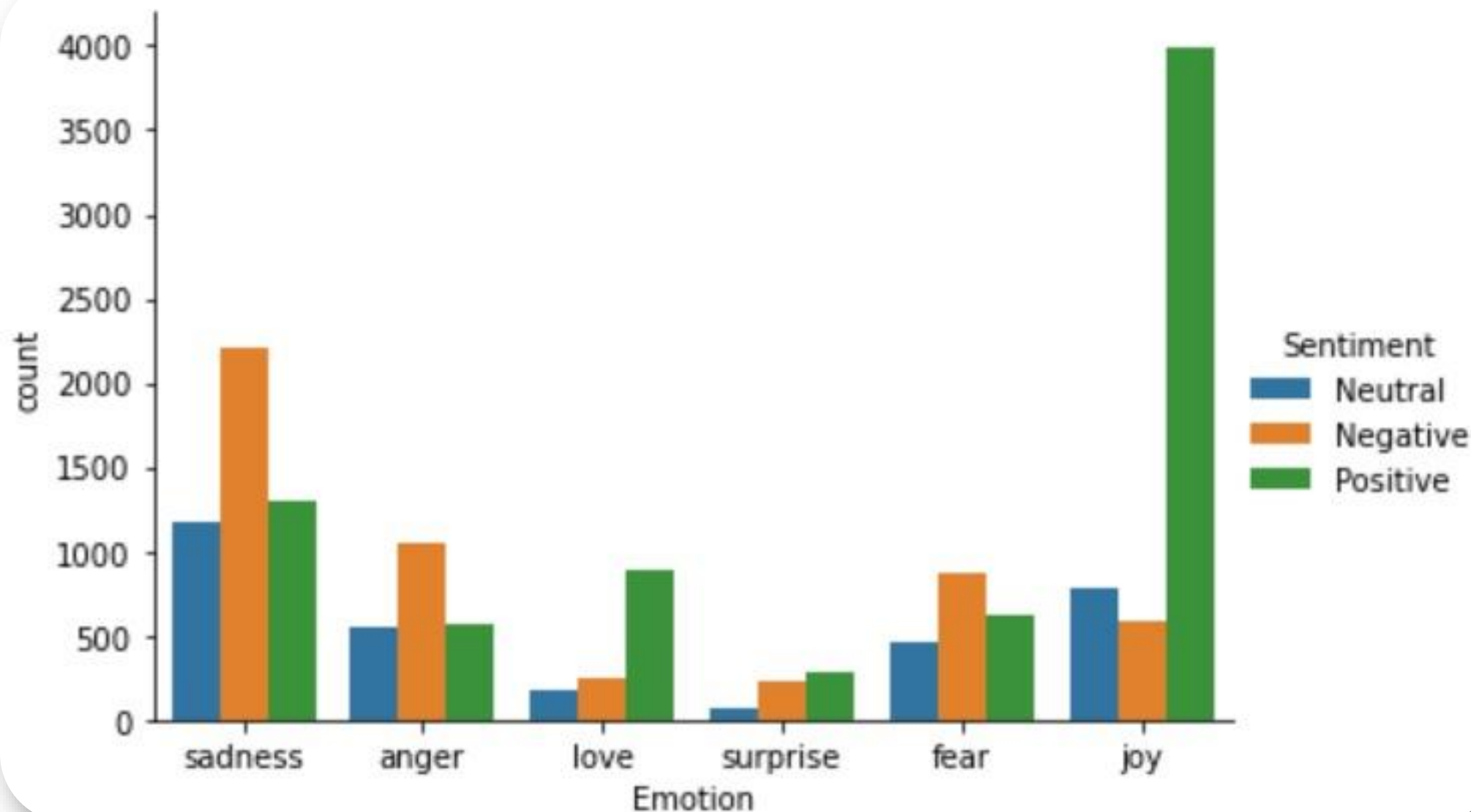


Análisis de sentimientos





Relación emociones y sentimientos



¿Cómo se analizan los datos?

01

"I didn't feel humiliated"
<<tokenización>>

02

[I didn't, feel humiliated]
<<Quitar contracciones>>

03

[I did, not, feel humiliated]
<<Quitar ruido>>

04

[feel, humiliated]
<<Normalización>>

05

[feel, humy, humiliated]



¿Por qué quitamos el like y el feel?

Text

Emotion

I **feel** like a faithful servant

Love

I am just **feeling** cranky and blue

Anger

I can have for a treat or if I am **feeling** fest...

Joy

I start to **feel** more appreciative of what g...

Joy

I for one am **feeling** a bit anxious at how I...

Fear

Text

Emotion

I **feel** incredibly lucky just to be able to tal...

Joy

I **feel** less keen about the army every day

Joy

I **feel** dirty and ashamed for saying that

Sadness

I **feel** bitchy but not defeates yet

Anger

Muestra al azar de 10 posts

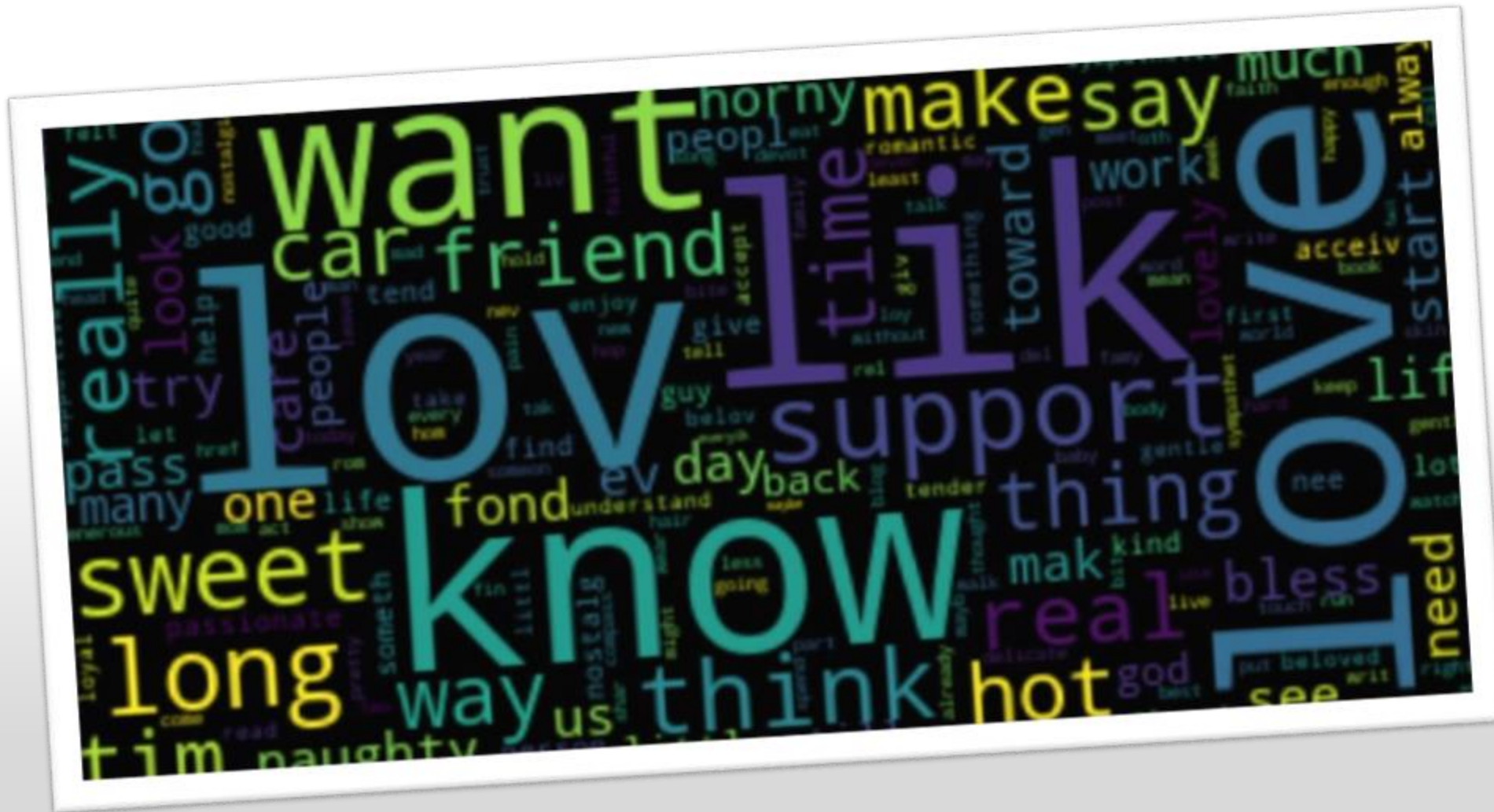
Text

Emotion

Text

Emotion

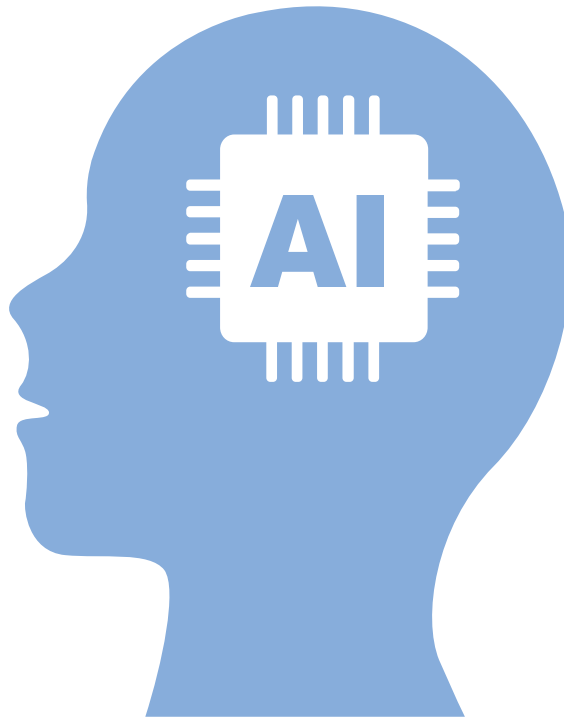
A Resultados de perfilamiento



Palabras clave:

- Love
- Sweet
- Want
- Hot
- Lov

Modelos de Clasificación



Trabajo de Minería de Textos

Realizaremos una tarea de aprendizaje supervisado

Usando como Técnica Clasificación Multiclase

Algoritmos a utilizar:

- Naive Bayes
- SVC (Support Vector Classification)
- OneVsRest

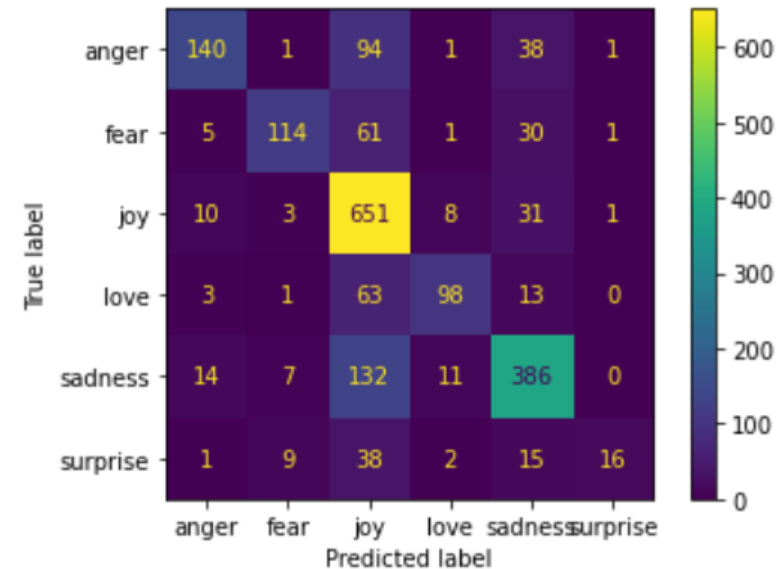
Modelo Escogido

OneVsRest Model

Métricas

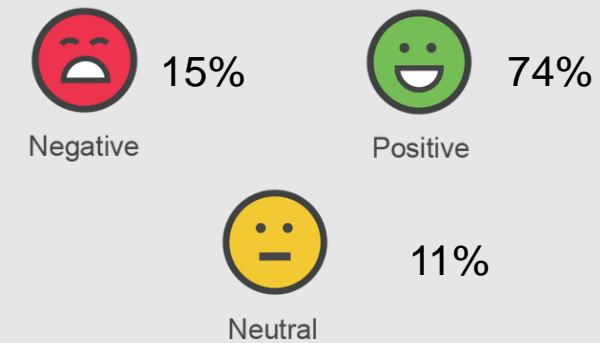
	precision	recall	f1-score	support
anger	0.81	0.51	0.62	275
fear	0.84	0.54	0.66	212
joy	0.63	0.92	0.75	704
love	0.81	0.55	0.66	178
sadness	0.75	0.70	0.73	550
surprise	0.84	0.20	0.32	81
accuracy			0.70	2000
macro avg	0.78	0.57	0.62	2000
weighted avg	0.73	0.70	0.69	2000

Matriz de confusión



Tablero de control Joy

Emotion	Positive	Neutral	Negative
joy	600	800	4000

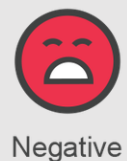
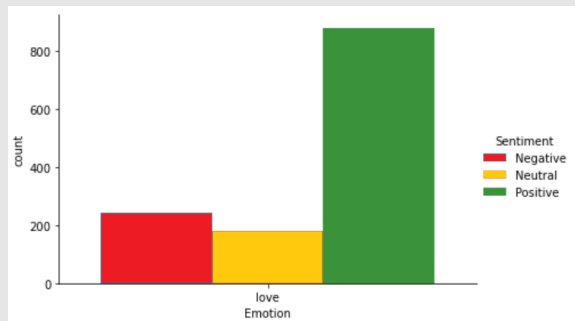


token	count
get	690
lik	590
go	480
make	470
want	450
know	450
think	420
would	410
mak	370
aser	360
good	350
tim	340
work	330
much	320
look	300
really	300
real	300
could	290
pretty	290
see	280
say	270
time	260
way	260
ev	250
happy	240
us	240
thing	230
back	230
love	210
help	210
enough	210
try	210
on	210
right	210
year	200
also	200
peopl	200
people	190
little	190
today	190
find	190
someth	180
lfr	180
lov	180
something	170
start	170
read	170
need	170
even	160
first	160



Tablero de control Love

Sentimientos



17%

Negative



68%

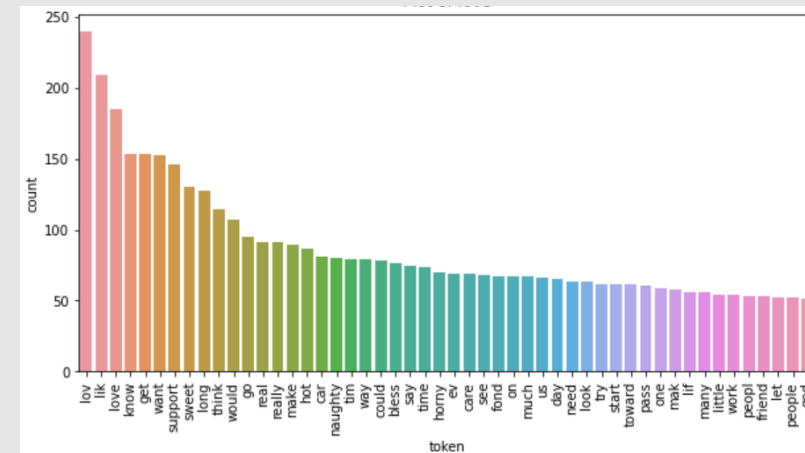
Positive



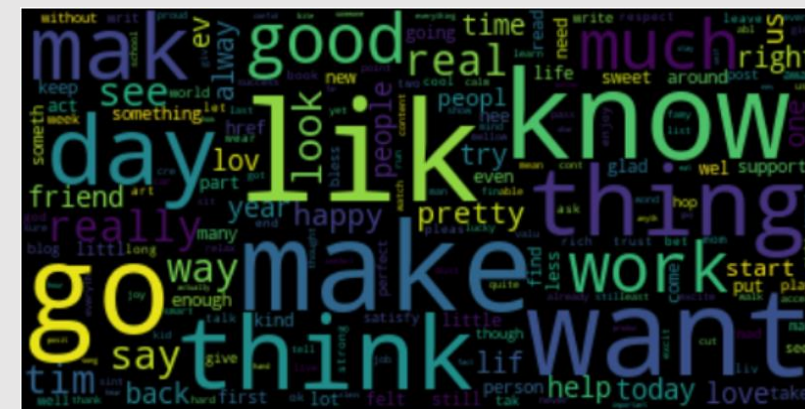
15%

Neutral

Palabras Clave



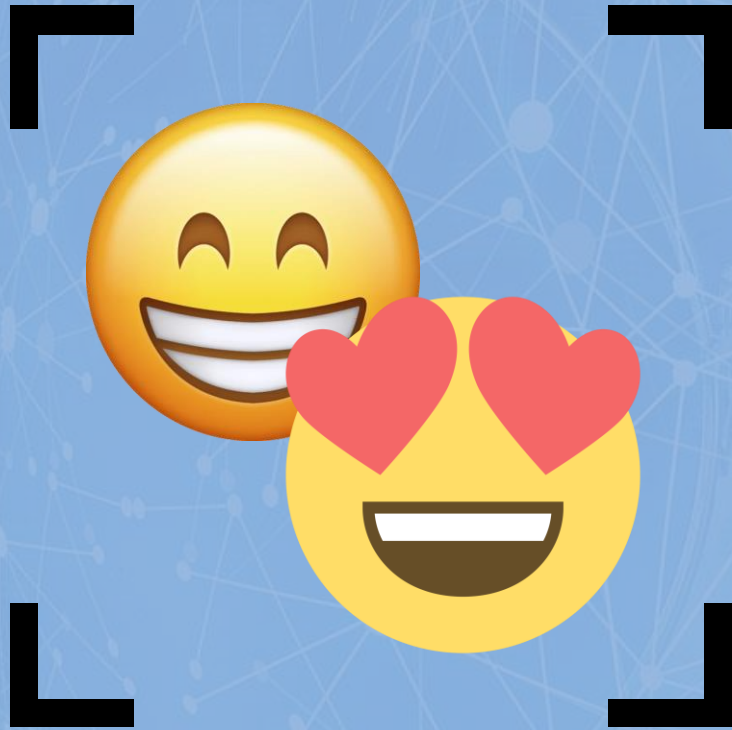
Nube de palabras



Conclusiones



Conclusiones



Emociones positivas: Love & Joy

- Identificar publicaciones que estén teniendo una buena reacción de los usuarios
- Criterio: Palabras clave de Joy y de Love
- Recopilar datos sobre estas publicaciones para poder entender el motivo que genera estas buenas reacciones
- Volver a realizar consultoría para que podamos identificar buenas prácticas que generan reacciones positivas



Thank You