

# Detección de palabras tóxicas



Alvarado Morán Óscar  
Bermúdez Marbán Dante  
García Avendaño Martín

IIMAS, UNAM.  
Licenciatura en Ciencia de Datos.



# Tabla de contenidos

- **Introducción**
  - Problemática
- **Corpus**
- **Metodología**
  - Modelos
- **Resultados**
- **Conclusiones y trabajo a futuro**



# Introducción: Problemática

La gran cantidad de textos ofensivos que la gente coloca en internet ocasionan problemas que evitan la creación de un ambiente saludable para todos los usuarios.

Existen sistemas que clasifican si un texto es tóxico o no pero no identifican qué parte, lo cual sería de gran ayuda para moderadores que tienen que revisar extensos comentarios.



# Corpus

Los datos a utilizar son una muestra aleatoria de 10K publicaciones que ya han sido catalogadas como ofensivas o como muy ofensivas

Del conjunto de 10K publicaciones se pidió a 3 grupos de personas que subrayaran los intervalos de palabras que resulten ofensivos

Si un texto contenía un mensaje ofensivo disfrazado de sarcasmo, que necesitase de una interpretación humana, que no fuera ofensivo o que todo el comentario fuera ofensivo, éste no debería tener ninguna anotación.

Trump is a bootlicking  
Russian traitor

What a shit show.



# Corpus

	spans	text
0	[8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19,...]	Another violent and aggressive immigrant killi...
1	[33, 34, 35, 36, 37, 38, 39]	I am 56 years old, I am not your fucking junio...
2	[0, 1, 2, 3]	Damn, a whole family. Sad indeed.
3	[7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]	What a knucklehead. How can anyone not know th...
4	[32, 33, 34, 35, 36, 37, 38]	"who do you think should do the killing?"\n\nA...
...	...	...
7934	[8, 9, 10, 11]	Another fool pipes in.
7935	[48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 5...	So if a restaurant owner puts up a sign saying...
7936	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,...]	Any faith that can't stand up to logic and rea...
7937	[5, 6, 7, 8, 9, 10, 11]	This idiotic. Use the surplus to pay down the ...
7938	[106, 107, 108, 109, 110, 169, 170, 171, 172, ...]	Who is this "we" of which you speak? Are you r...

spans	text
[8, 9, 10, 11]	Another fool pipes in.



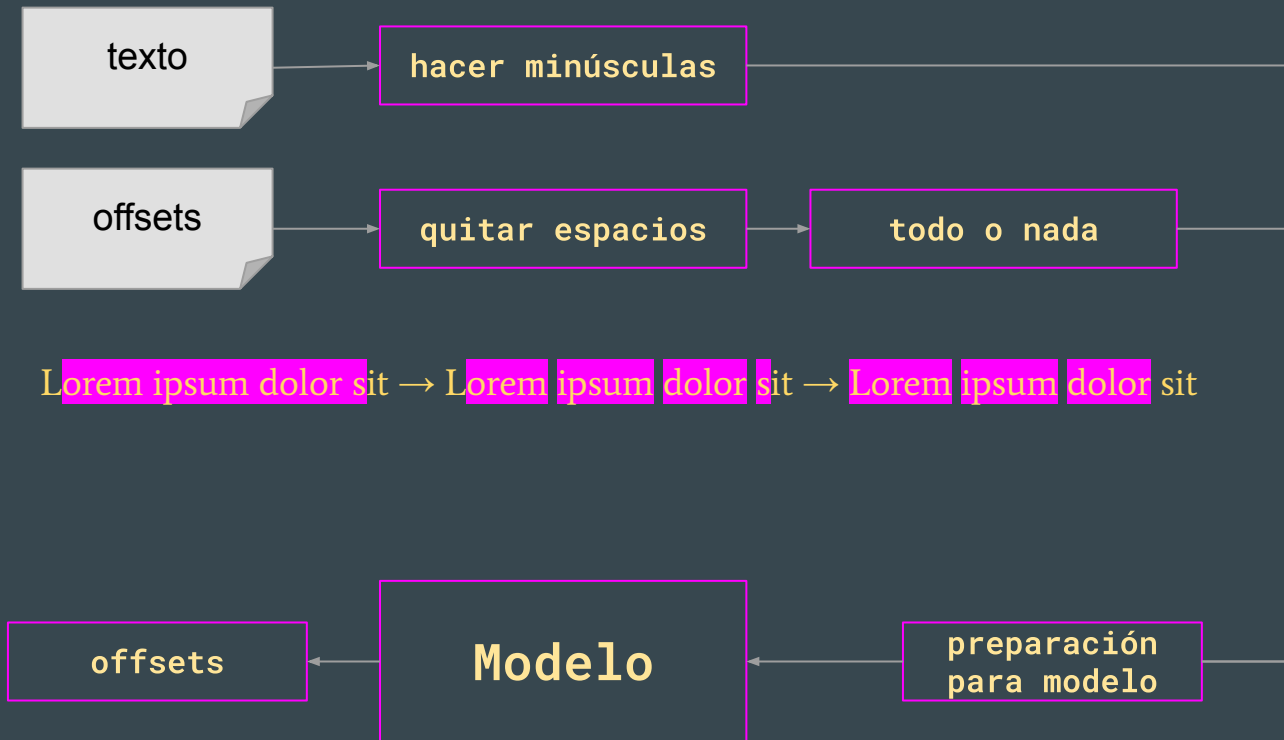
A n o t h e r f o o l p i p e s i n .  
□ □ □ □ □ □ □ □ 8 9 10 11 □ □ □ □ □ □ □ □ □ □

Any faith that can't stand up to  
logic and reason deserves to die.

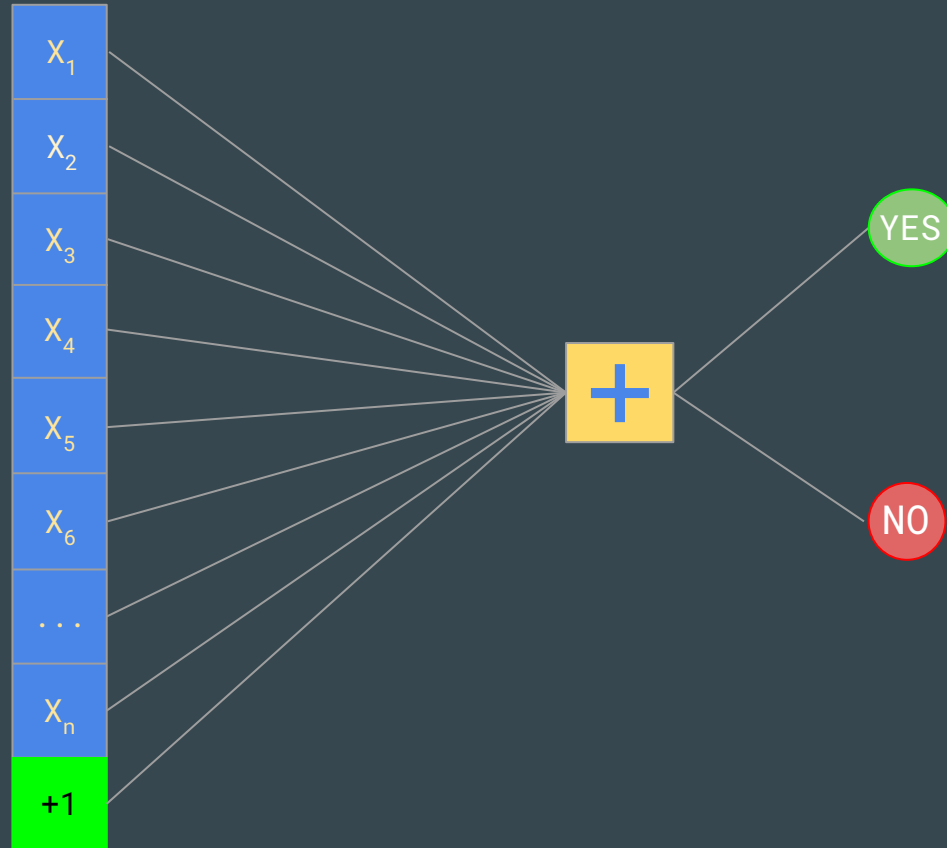
[]



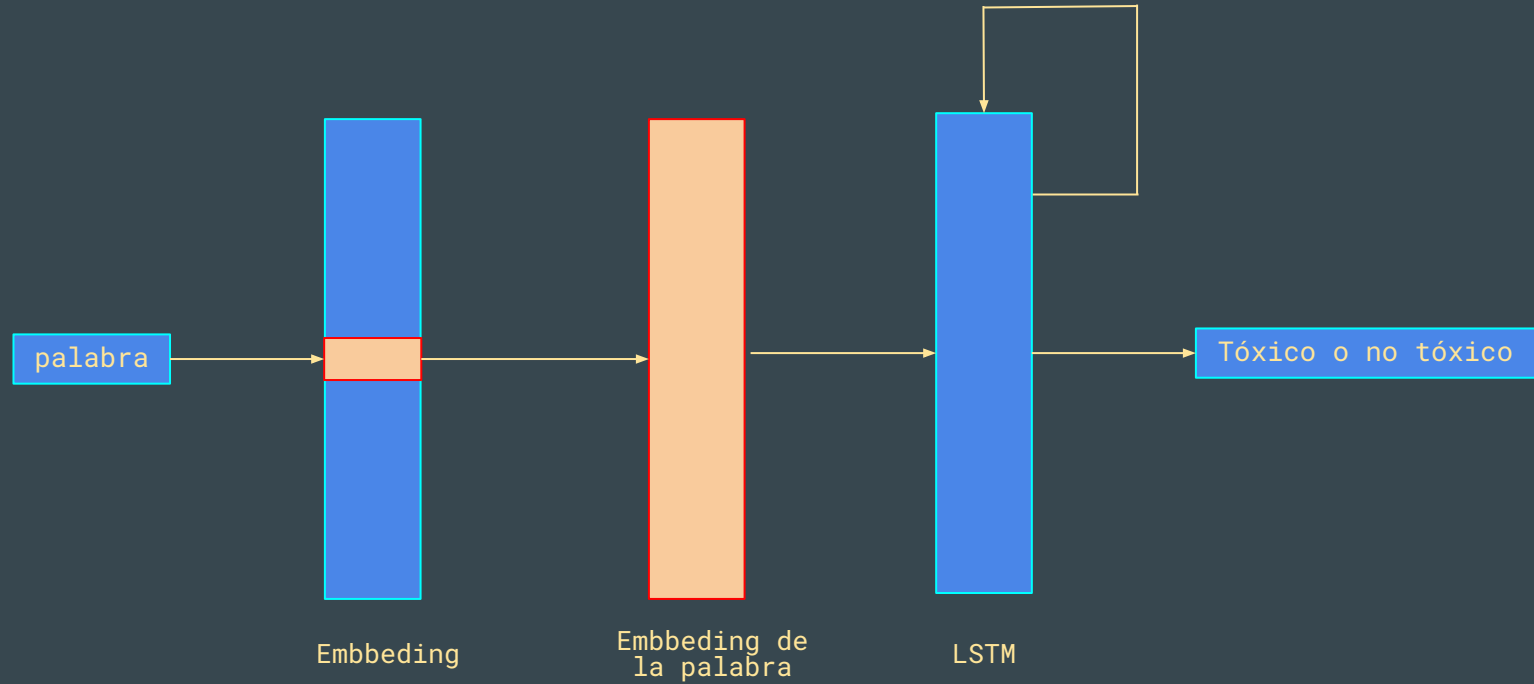
# Metodología



# Perceptrón



# Red neuronal





# Resultados

	Train	Test
Modelo oculto de Markov (binario)	0.585	0.284
Modelo oculto de Markov (con POS)	0.660	0.310
Perceptrón (binario)	0.933	0.490
Perceptrón (con pos)	0.890	0.491
Red neuronal	0.604	0.437



# Conclusiones y trabajo a futuro

- Mejor modelo: perceptrón
- Pésima calidad de subrayado
  - Errores humanos incorregibles
- Posible mejora con posprocesamiento
  - Considerar espacios
  - Tomar en cuenta toda la oración o no tomarla en cuenta

