

Galileo
UNIVERSIDAD
La Revolución en la Educación



Universidad Galileo
Instituto de Investigación de Operaciones
Posgrado en Análisis y Predicción de Datos

Antonio Navas
Daniel Rodríguez
Luis Florian

Agenda

- Objetivos
- Origen y estructura del Dataset utilizado
- Nombre de Películas
- Cantidad de Diálogos
- Normalización
- TF-IDF
- Palabras más frecuentes
- Análisis de Sentimientos
- Conclusiones



Objetivo

- Explorar y analizar los diálogos de la trilogía **The Lord of The Rings**.
- Realizar análisis de sentimientos por medio de técnicas de Text Mining.
- Graficar y analizar la cantidad de palabras con emociones de positivismo, negativismo, miedo, confianza, etc.
- Comparar resultados con más de una herramienta de análisis de sentimientos.

Análisis de Sentimientos

El análisis de sentimientos es el proceso de determinar "computacionalmente" si un escrito es positivo, negativo o neutral. También se conoce como minería de opiniones, derivar la opinión o actitud de un hablante.

El análisis de sentimiento intenta determinar la actitud de un interlocutor o usuario con respecto a algún tema o la polaridad contextual general de un documento.

Dataset

Trabajamos con un archivo en formato CSV el cual contiene los diálogos de la trilogía, donde cada fila contiene la información del personaje al que pertenece el diálogo y el nombre de la película

```
lotr_df = pd.read_csv("lotr_scripts.csv")  
lotr_df.head(10)
```



Unnamed: 0	char	dialog	movie
0	DEAGOL	Oh Smeagol Ive got one! , Ive got a fish Smeag...	The Return of the King
1	SMEAGOL	Pull it in! Go on, go on, go on, pull it in!	The Return of the King
2	DEAGOL	Arrghh!	The Return of the King
3	SMEAGOL	Deagol!	The Return of the King
4	SMEAGOL	Deagol!	The Return of the King
5	SMEAGOL	Deagol!	The Return of the King
6	SMEAGOL	Give us that! Deagol my love	The Return of the King
7	DEAGOL	Why?	The Return of the King
8	SMEAGOL	Because' , it's my birthday and I wants it.	The Return of the King
9	SMEAGOL	My precious.	The Return of the King

https://www.kaggle.com/paultimothymooney/lord-of-the-rings-data?select=lotr_scripts.csv

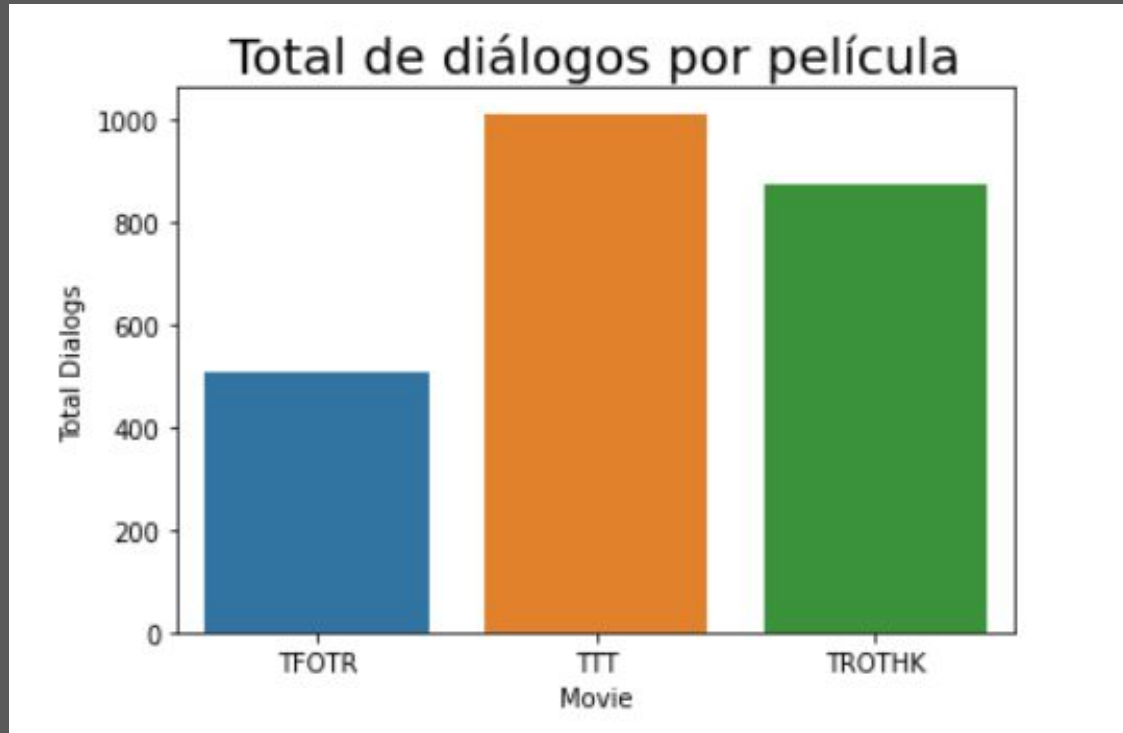
Nombre de las Películas

- The Fellowship of the Ring (TFOTR)
- The Two Towers (TTT)
- The Return of The King (TROTHK)

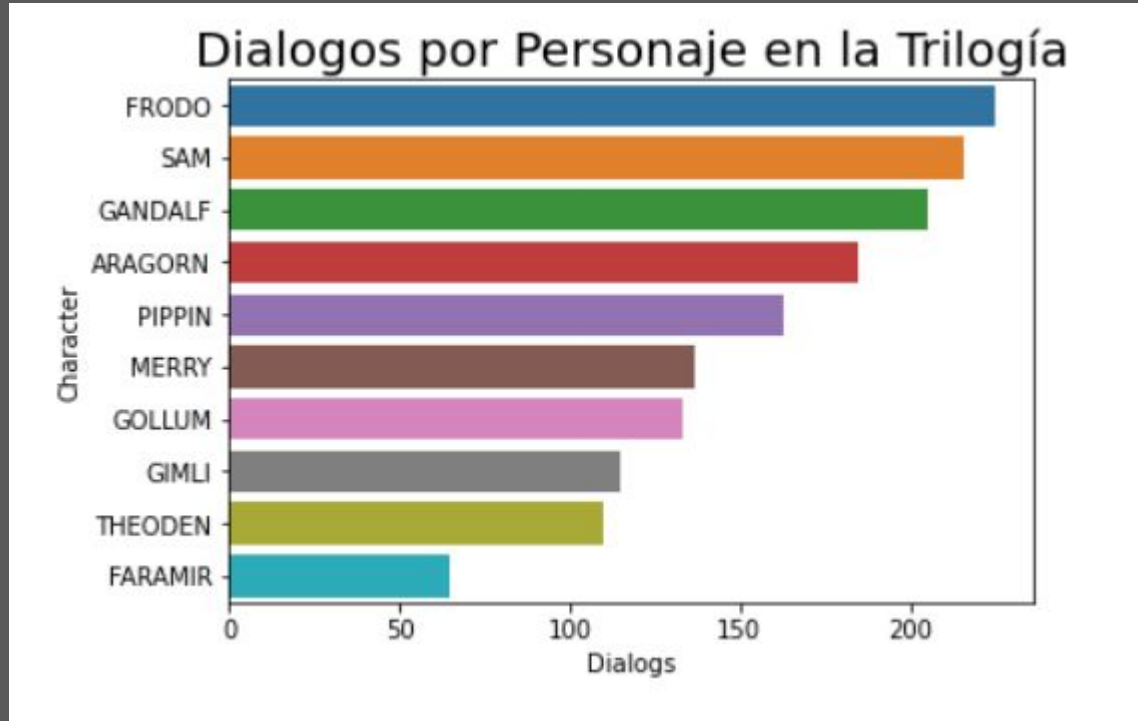
```
movies_name = np.unique(lotr_df['movie'])  
print(movies_name)
```

```
['The Fellowship of the Ring ' 'The Return of the King ' 'The Two Towers ']
```

Diálogos

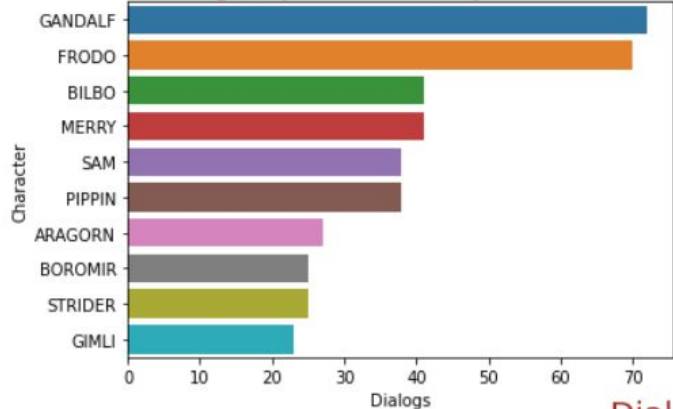


Diálogos por Personaje

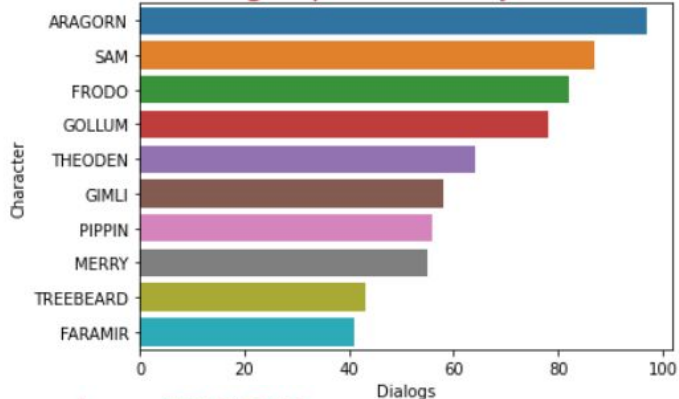


Diálogos por Personaje

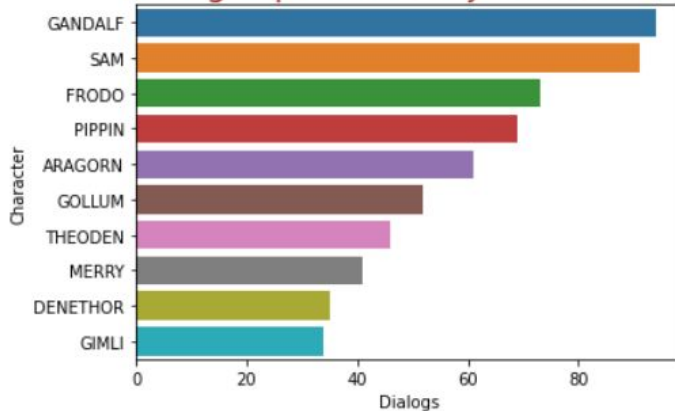
Dialogos por Personaje en TFOTR



Dialogos por Personaje en TTT



Dialogos por Personaje en TROTHK



Normalización

Metodología:

- Remover Caracteres Especiales
- Tokenización de Palabras
- Remover Stopwords
- Stemming
- Lematización
- Crear Bag of Words

	char	clean_dialogs
0	DEAGOL	oh smeagol got one got fish smeagol smeagol
1	SMEAGOL	pull go go go pull
2	DEAGOL	arrghh
3	SMEAGOL	deagol
4	SMEAGOL	deagol
5	SMEAGOL	deagol
6	SMEAGOL	give u deagol love
7	DEAGOL	
8	SMEAGOL	birthday want
9	SMEAGOL	preciou

TF-IDF

Listado de las palabras no comunes

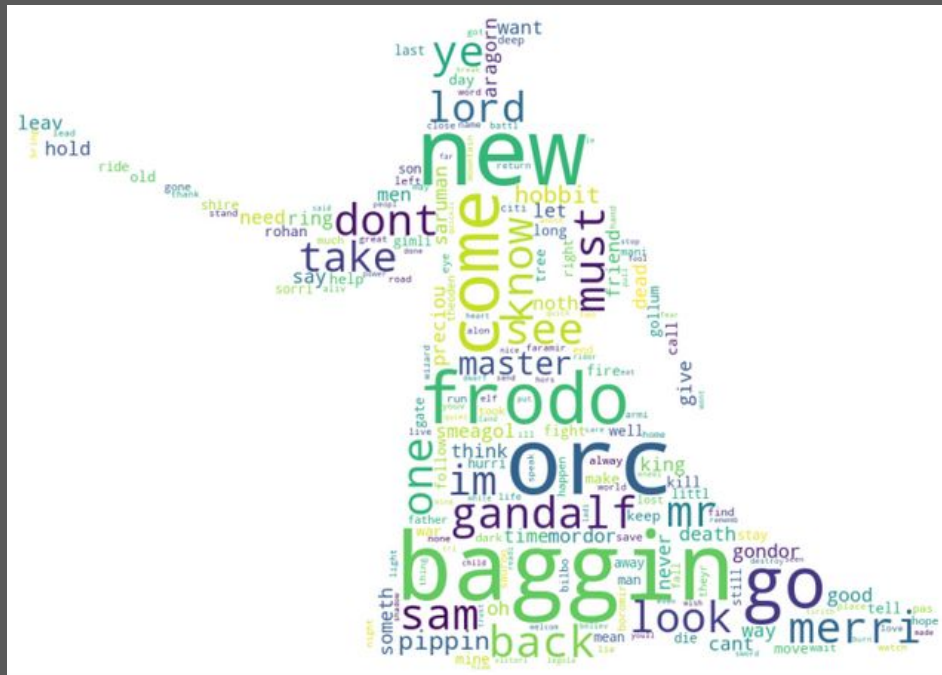
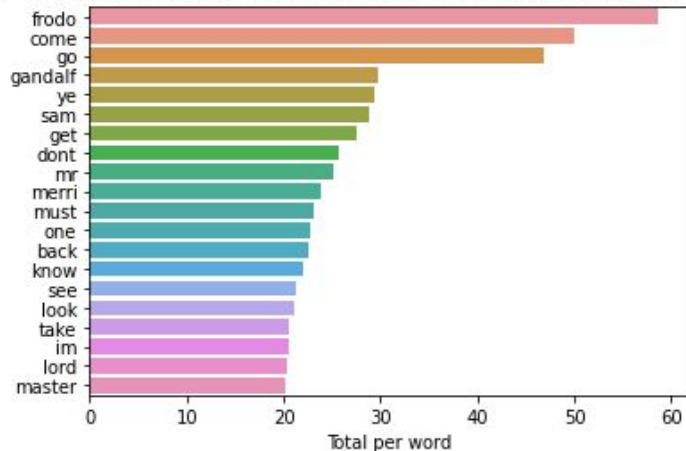
TF-IDF max per document	
0	smeagol
1	pull
2	arrghh
3	deagol
4	deagol
...	...
2383	merri
2384	merri
2385	alway
2386	hardi
2387	foolhardi

2388 rows × 1 columns

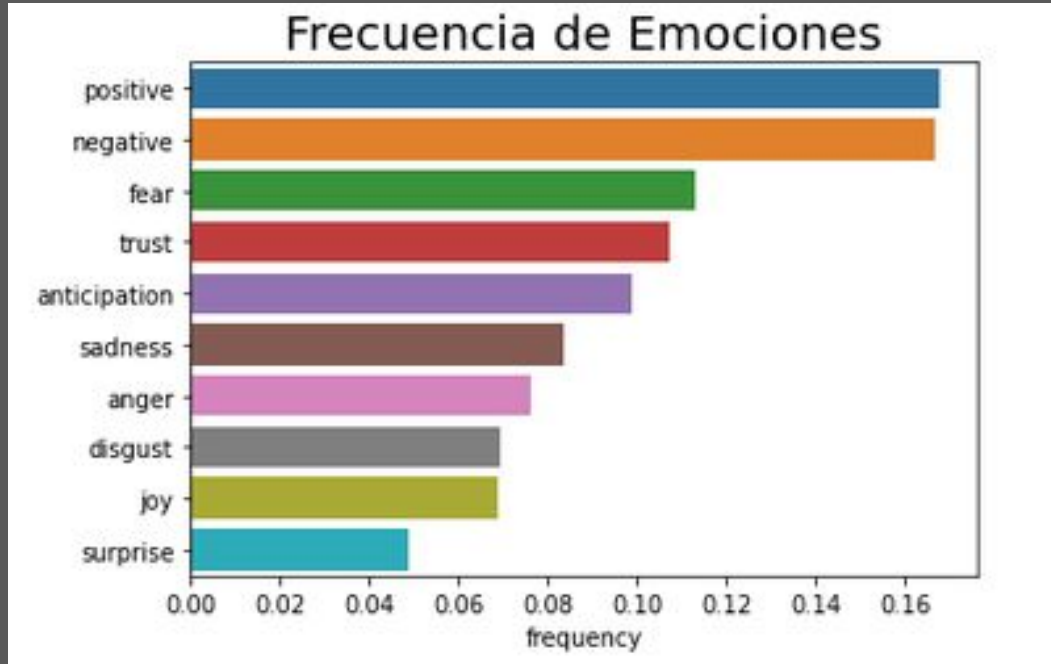
```
tfidf_table.shape  
  
(2388, 2644)
```

Palabras más frecuentes

Palabras más frecuentes en todo el Corpus



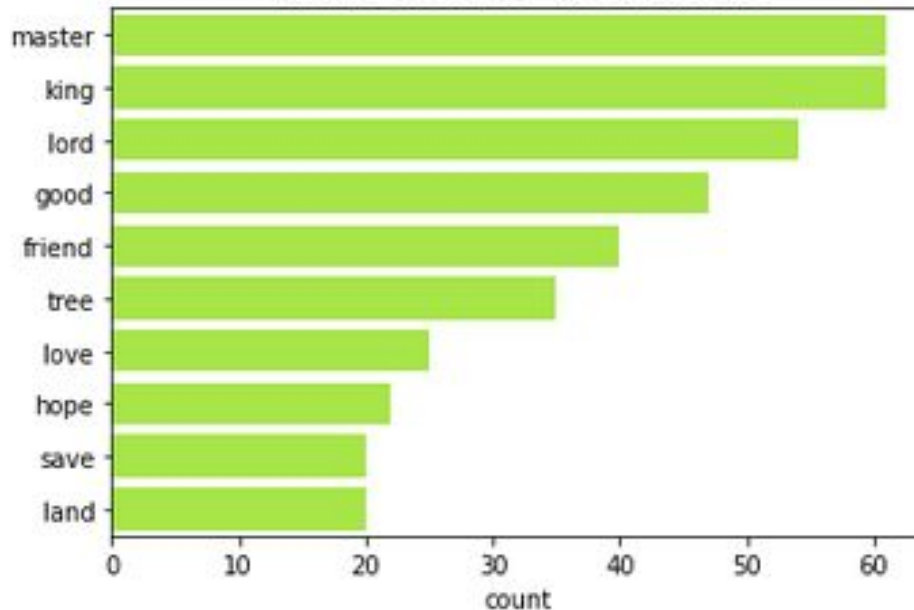
Análisis de Sentimiento (NRCLex)



Frecuencia de Emociones	
	frequency
positive	0.167925
negative	0.166538
fear	0.112926
trust	0.107071
anticipation	0.098752
sadness	0.083808
anger	0.076105
disgust	0.069327
joy	0.068556
surprise	0.048991

Positismo

Emociones Positivas

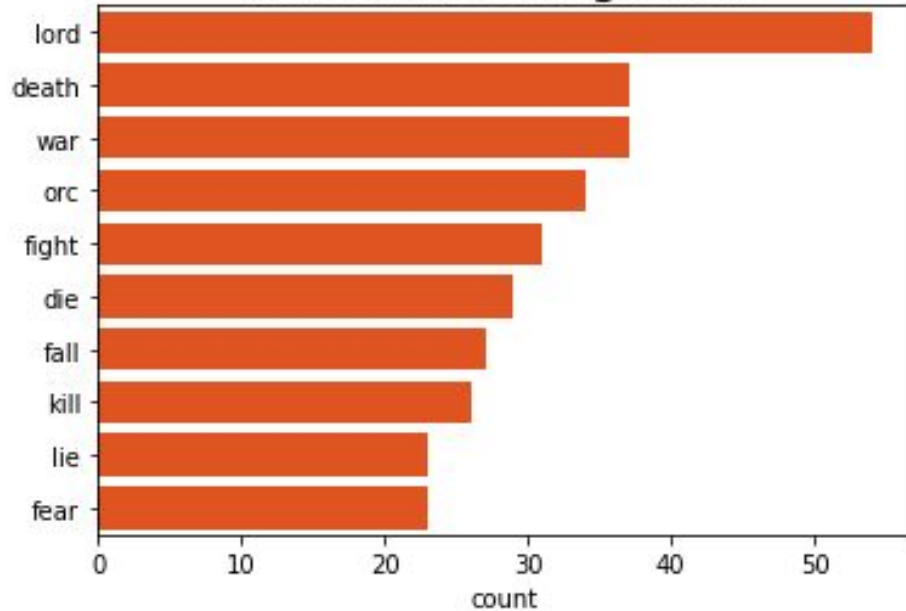


Top 10 de Palabras Positivas

	count
master	61
king	61
lord	54
good	47
friend	40
tree	35
love	25
hope	22
save	20
land	20

Negativismo

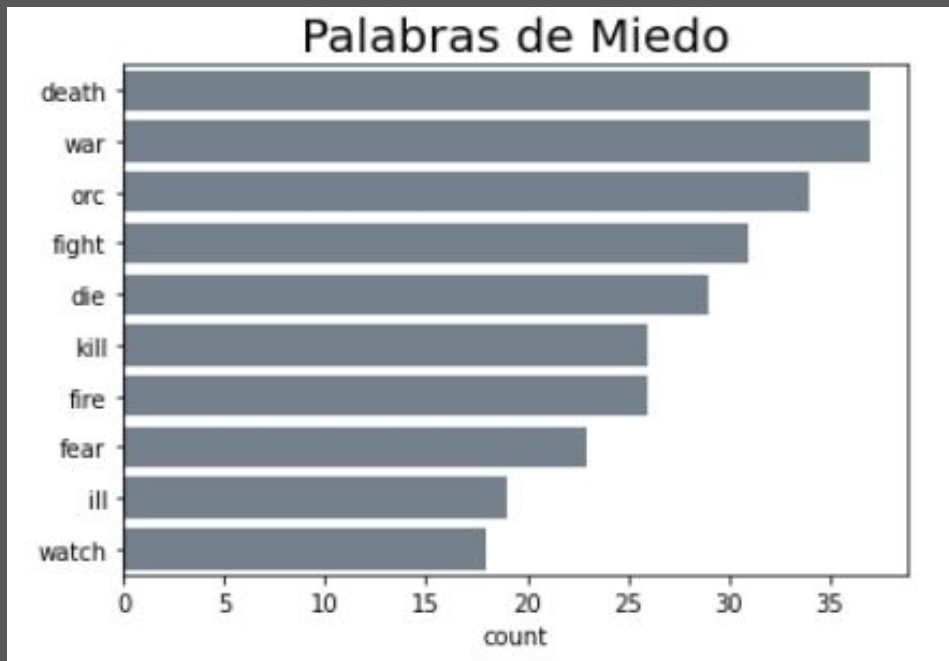
Emociones Negativas



Top 10 Palabras Negativas

	count
lord	54
death	37
war	37
orc	34
fight	31
die	29
fall	27
kill	26
lie	23
fear	23

Miedo

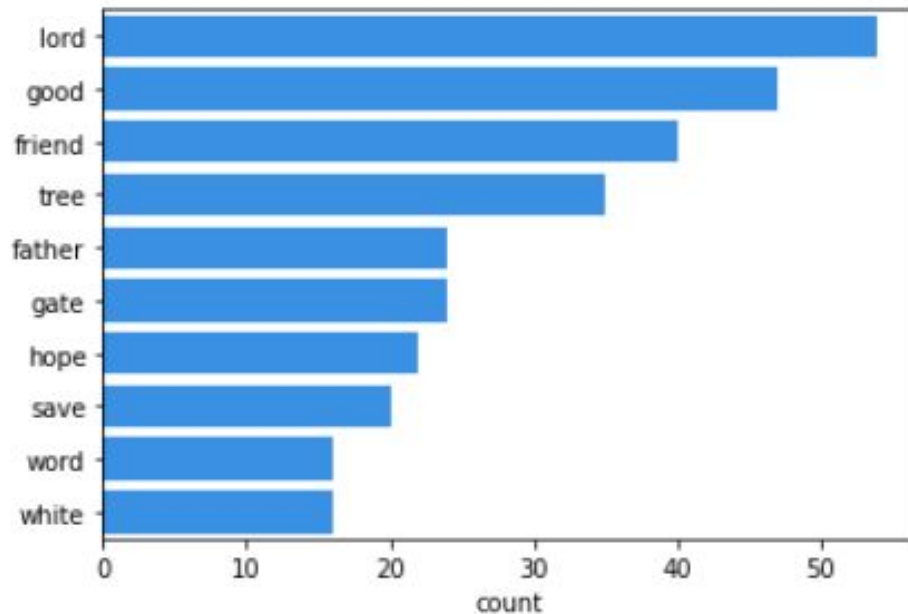


Top 10

	count
death	37
war	37
orc	34
fight	31
die	29
kill	26
fire	26
fear	23
ill	19
watch	18

Confianza

Palabras de Confianza

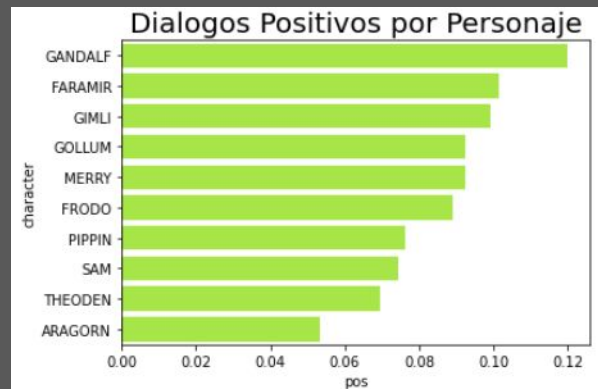


Top 10 Palabras de Confianza

	count
lord	54
good	47
friend	40
tree	35
father	24
gate	24
hope	22
save	20
word	16
white	16

Análisis de Sentimiento en los diálogos (Vader)

	character	neg	neu	pos
0	FRODO	0.045676	0.860707	0.089169
1	SAM	0.106194	0.819310	0.074491
2	GANDALF	0.111995	0.768102	0.119898
3	ARAGORN	0.079243	0.867389	0.053357
4	PIPPIN	0.072319	0.851233	0.076436
5	MERRY	0.041409	0.866036	0.092577
6	GOLLUM	0.153188	0.746699	0.092586
7	GIMLI	0.115026	0.785670	0.099322
8	THEODEN	0.117600	0.812800	0.069618
9	FARAMIR	0.095092	0.803292	0.101600



Conclusiones

- El análisis de sentimiento nos puede ayudar a determinar el tono emocional detrás de una serie de palabras.
- Nos ayuda a comprender el estado de ánimo de los personajes o cliente, según sea el caso.
- La ingeniería característica como en otras aplicaciones, es de vital importancia para el valor de los resultados.

FIN

Pippin: Nunca pensé en este final

Gandalf: ¿Final? No, el viaje no concluye aquí. La muerte es sólo otro sendero, que recorreremos todos. El velo gris de este mundo se levanta y todo se convierte en plateado cristal. Es entonces, cuando se ve...

Pippin: ¿Qué, Gandalf? ¿Qué se ve?

Gandalf: Un 100 en nuestro proyecto de Text Mining

