

# Minería de Texto en R

## Escuela de Métodos 2026

26 enero 2026  
Dr. Enrique García Tejeda

Hola,

Soy Sir Leonard Valentinovich Blavatnik, un empresario británico, inversionista, filántropo y presidente (Access Industries).

Di destinó el 20 por ciento de mi riqueza personal a obras de caridad. Y también prometí

para regalar el resto del 20% este año 2023 a Particulares.. Tengo decidido donarle

320.000,00 € (trescientos veinte mil euros).

Si usted es interesado en mi donación,

póngase en contacto conmigo para obtener más información. También puedes leer más sobre mí a través del siguiente enlace

[https://en.m.wikipedia.org/  
wiki/Len\\_Blavatnik](https://en.m.wikipedia.org/wiki/Len_Blavatnik)

# La estafa del príncipe nigeriano



Hola,

Soy Sir Leonard Valentinovich Blavatnik, un empresario británico, inversionista, filántropo y presidente (Access Industries). Di destinó el 20 por ciento de mi riqueza personal a obras de caridad. Y también prometí para regalar el resto del 20% este año 2023 a Particulares.. Tengo decidido donarle 320.000,00 € (trescientos veinte mil euros). Si usted es interesado en mi donación, póngase en contacto conmigo para obtener más información. También puedes leer más sobre mí a través del siguiente enlace [https://en.m.wikipedia.org/wiki/Len\\_Blavatnik](https://en.m.wikipedia.org/wiki/Len_Blavatnik)

## La estafa del príncipe nigeriano



El correo basura es una muestra de un caso de éxito de procesamiento estadístico e informático del lenguaje natural.

1. Convierte la estructura del texto en una matriz estadística.
2. Asigna frecuencias a las palabras y estima correlaciones
3. Distingue los correos etiquetados por humanos como basura vs correos deseados
4. Halla las palabras más presentes en los correos etiquetados como basura
5. Con un nuevo correo, el algoritmo etiqueta como correo basura o correo deseado el texto entrante.
6. El algoritmo aprende a identificar un correo no deseado con entrenamiento

## La estafa del príncipe nigeriano



El correo basura es una muestra de un caso de éxito de procesamiento estadístico e informático del lenguaje natural.

1. Convierte la estructura del texto en una matriz estadística.
2. Asigna frecuencias a las palabras y estima correlaciones
3. Distingue los correos **etiquetados** por humanos como basura vs correos deseados
4. Halla las palabras más presentes en los correos etiquetados como basura
5. Con un nuevo correo, el algoritmo etiqueta como correo basura o correo deseado el texto entrante.
6. El algoritmo **aprende a identificar** un correo no deseado con entrenamiento

## La estafa del príncipe nigeriano



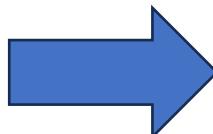
El correo basura es una muestra de un caso de éxito de procesamiento estadístico e informático del lenguaje natural.

1. Convierte la estructura del texto en una matriz estadística.
2. Asigna frecuencias a las palabras y estima correlaciones
3. Distingue los correos **clasificados** por humanos como basura vs correos deseados
4. Halla las palabras más presentes en los correos etiquetados como basura
5. Con un nuevo correo, el algoritmo etiqueta como correo basura o correo deseado el texto entrante.
6. El algoritmo **predice** si un correo es no deseado con entrenamiento

## Textos como matrices

Correo	príncipe	ayuda	donación	amigo	Spam
1º	4	2	1	2	Si
2º	0	1	0	4	No
3º	2	2	2	1	Si
4º	0	1	1	0	Si
X	0	3	2	1	???

# Textos como matrices



Correo	príncipe	ayuda	donació	amigo	Spam
1º	4	2	1	2	Si
2º	0	1	0	4	No
3º	2	2	2	1	Si
4º	0	1	1	0	Si
X	0	3	2	1	???

El auge de los algoritmos de aprendizaje de máquina proviene de la gran cantidad actual de datos:

1. Volumen
2. Velocidad
3. Variedad

Las 3 v's del Big data

1 DATOS



2 LIMPIOS EN UNA BASE DE DATOS



3 ANALIZADOS



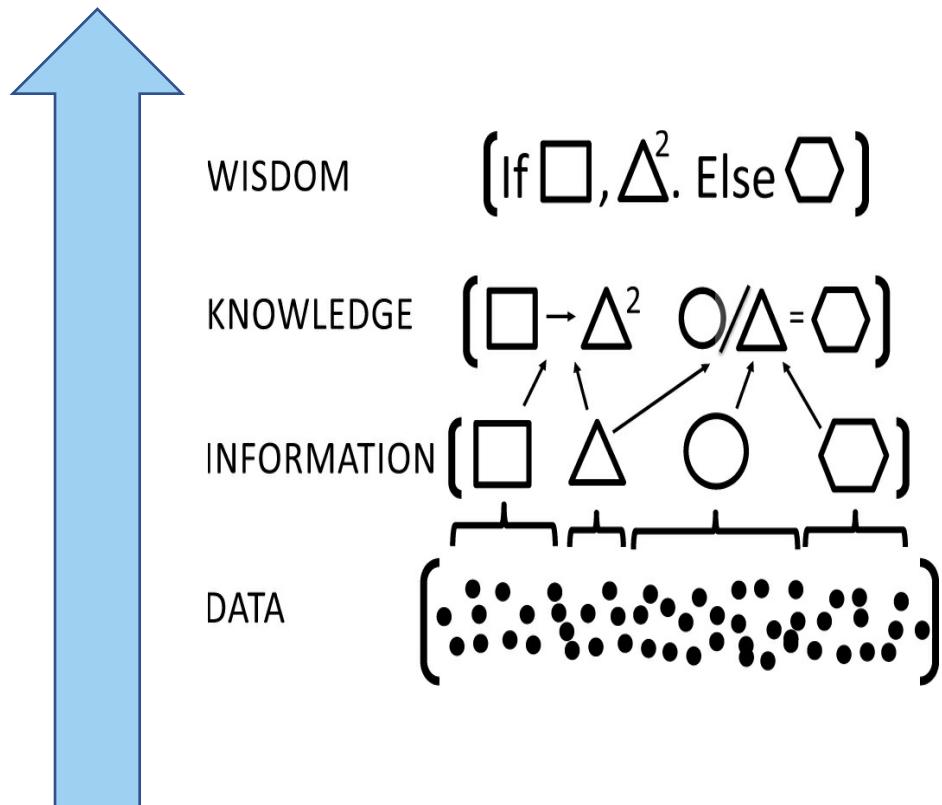
4 PRESENTADOS DE FORMA VISUAL



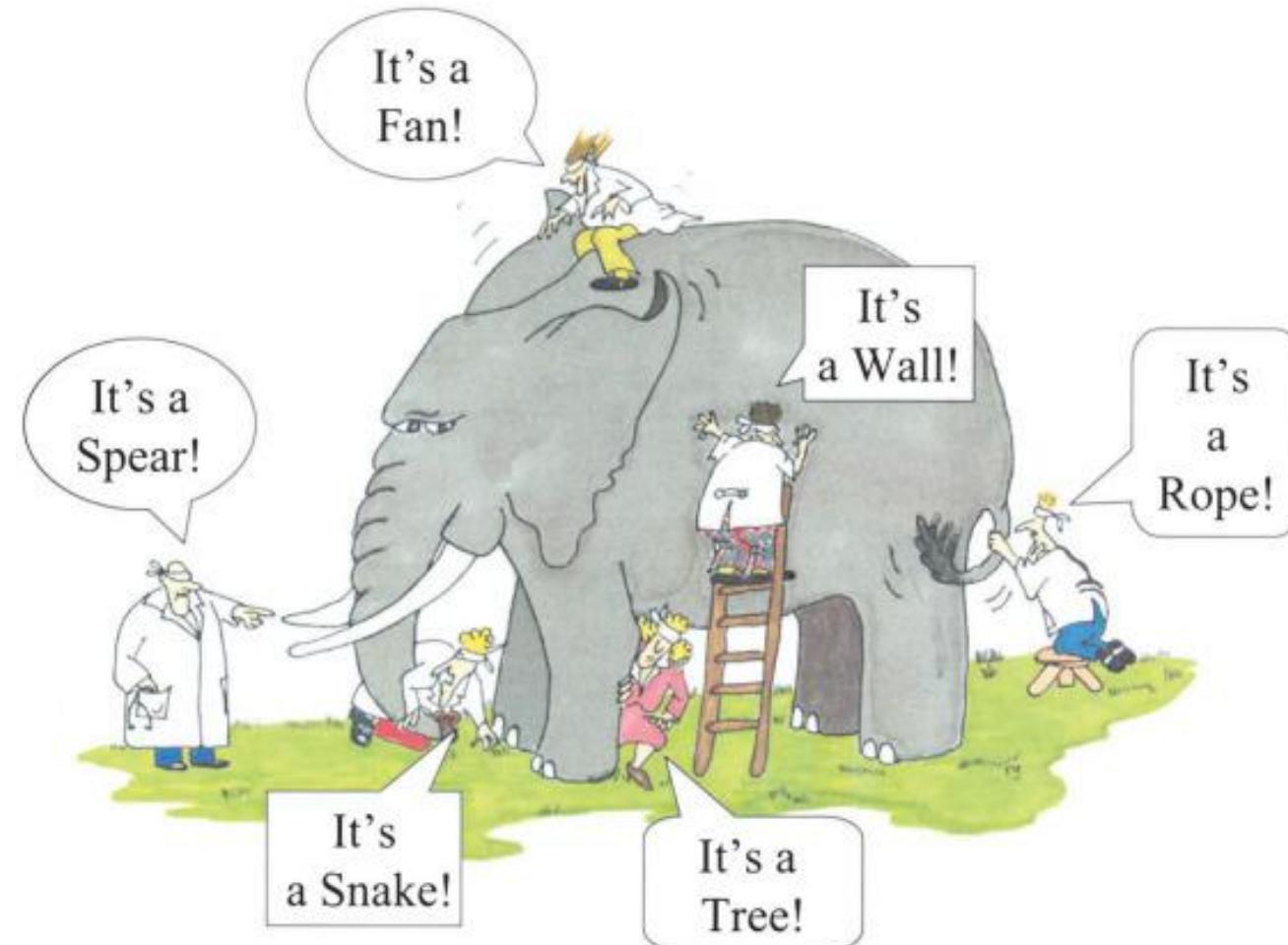
5 EXPLICADOS CON UNA HISTORIA



# Datos instantáneos, contingentes y desorganizados



# Las limitaciones del análisis de datos



**Fuente:** Daigneault, P. M. (2013). The blind men and the elephant: A metaphor to illuminate the role of researchers and reviewers in social science. *Methodological innovations online*, 8(2), 82-89.

# Minería de Texto en R

## Escuela de Métodos 2026

26 enero 2026  
Dr. Enrique García Tejeda