

# ConectaR 2021

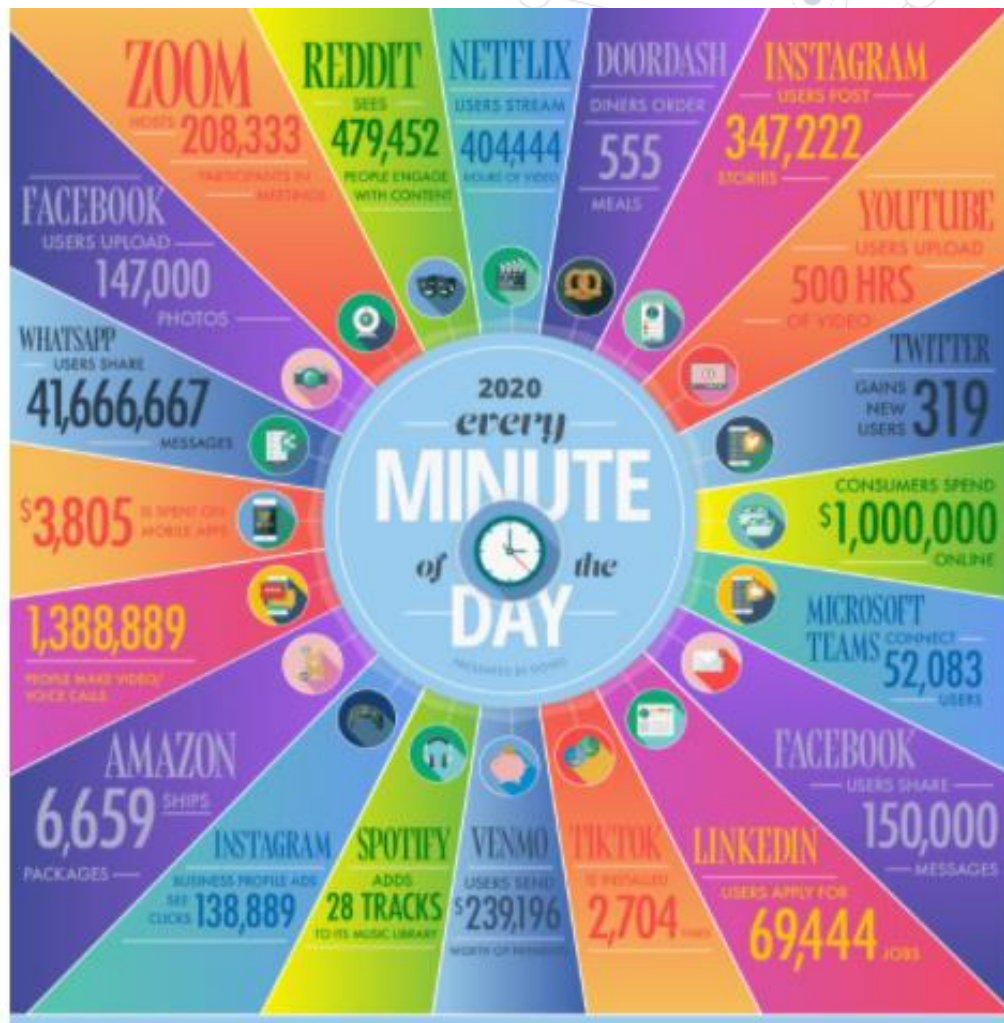
## • **Análisis de Sentimientos** • **con Rstudio**



# CONTEXTO

Durante ese periodo de tiempo, Twitter añade 319 usuarios nuevos y eso explica que la red social haya experimentado el mayor crecimiento en número de usuarios de su historia en el último año y que a YouTube se suben más de 500 horas de vídeo.

Infografía : Domo



## ¿ Que es el Análisis de Sentimientos?

El análisis de sentimientos también conocido como minería de opinión (opinion mining), es el estudio sistemático que interpreta las actitudes, opiniones y sentimiento del usuario hacia una marca. Consiste en determinar el tono emocional que hay detrás de una serie de palabras y clasificarlas como positiva, negativa o neutra.



## ¿ Que es el Análisis de Sentimientos?

Hay un notable interés y necesidad por parte de compañías, políticos, así como consumidores, de contar con herramientas que les permitan identificar y sintetizar las opiniones y sentimientos que las personas expresan en blogs, reseñas de productos, redes sociales y en general cualquier tipo de repositorio digital.



## ¿ Que preguntas responde?

¿Una reseña de producto determinada es positiva o negativa?

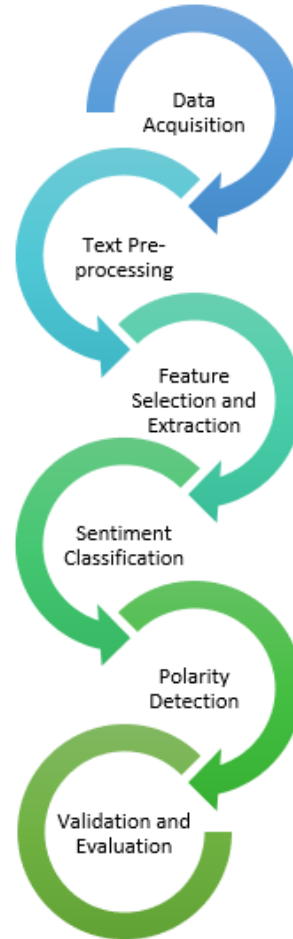
¿Un cliente está satisfecho o insatisfecho en función de su respuesta por correo electrónico?

Según una muestra de tweets, ¿cómo responde la gente a una determinada campaña publicitaria, lanzamiento de producto o noticia?

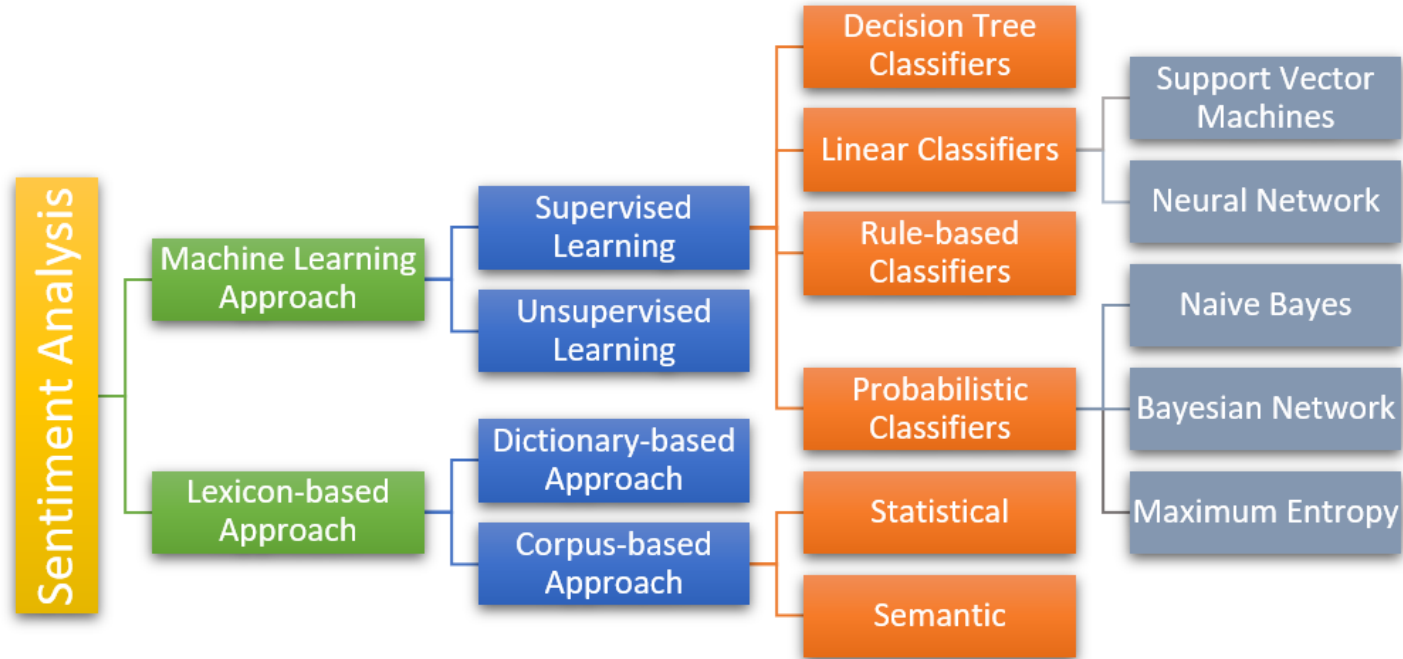
¿Cómo han cambiado las actitudes de los blogueros sobre el presidente desde las elecciones?

¿Cómo han cambiado las actitudes de los tuiteros desde la Presidencia de Merino (Perú)?

# Proceso de Análisis de Sentimientos



# Enfoques para Análisis de Sentimientos



# ¿Cuáles son las ventajas y limitaciones de los enfoques?

## Basado en aprendizaje automático

Ventaja : A diferencia de los basados en léxico, estos modelos se pueden construir para un propósito o contexto específico.

Limitación : La obtención de datos etiquetados para la capacitación puede resultar difícil o costosa.

## Basado en léxico

Ventaja : No se requiere formación.

Limitación : la precisión depende de los recursos léxicos. Número finito de palabras en léxicos y asignación de una orientación de sentimiento fija y puntuación a las palabras.





# En Rstudio

dplyr (Wickham y Francois [2016](#) )

tidyr (Wickham [2016](#) )

ggplot2 (Wickham [2009](#) )

broom (Robinson [2017](#) )

Al mantener la entrada y la salida en tablas ordenadas, los usuarios pueden realizar una transición fluida entre estos paquetes.

Estas herramientas ordenadas se extienden naturalmente a muchos análisis y exploraciones de texto.

# Definiciones Claves

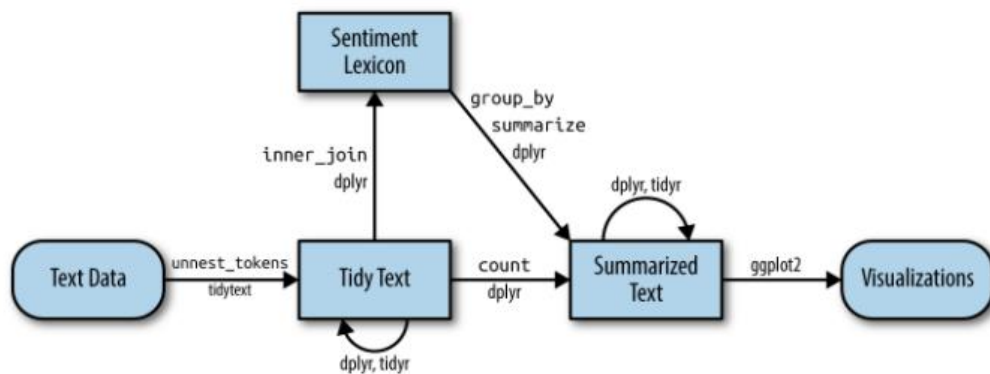
## Formato de Texto Ordenado

Permiten que el manejo de datos sea más fácil y efectivo. Tienen una estructura específica :

Cada variable es una columna, cada observación una fila, cada tipo de unidad de observación es una tabla.

## Análisis de sentimiento con datos ordenados

Se aborda la comprensión de la intención emocional de las palabras para inferir si una sección de texto es positiva ó negativa ó aborda otra emoción como sorpresa ó disgusto.



## En Rstudio

El paquete tidytext contiene 3 diccionarios distintos:

**AFINN:** asigna a cada palabra un valor entre -5 y 5. Siendo -5 el máximo de negatividad y +5 el máximo de positividad.

**bing:** clasifica las palabras de forma binaria positivo/negativo.

**nrc:** clasifica cada palabra en uno o más de los siguientes sentimientos: positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.



**Vamos a  
Rstudio!**

# ANÁLISIS DE SENTIMIENTO CON TWITTER



Caso :  
tweets!



Para obtener los sentimientos relacionados a los comentarios de dos tendencias de Twitter en el contexto de la vacancia presidencial.

- Twitter es actualmente una dinámica fuente de contenidos
- Dada su popularidad e impacto, se ha convertido en uno de los principales medios de difusión de los principales medios de comunicación tradicionales (radio y televisión).

# PROCESO DE EXTRACCIÓN DE DATOS

1. Acceder a una cuenta de twitter.
2. Crear una nueva App en: <https://apps.twitter.com/>
3. Conseguir tus claves en la pestaña: Keys and Access Tokens
4. Finalmente ya podemos usarlas en el paquete de R.

## LIBRERÍAS

1. `library(twitterR)` : para el uso de las 4 claves de la Api de Twitter
2. `library(rtweet)`: Recopilar datos de twitter
3. `library(nrc)`: clasificar palabras en emociones
4. `library(ggplot2)`: gráficos de barras

# PROCESO DE EXTRACCIÓN DE DATOS

Ejemplo :

Tweets – Noviembre 2020

Candidatos

Julio Guzman

George Forsyth

Cesar Acuña



## CONTEXTO Y CONTENIDO

## Vacancia presidencial a Martín Vizcarra

Un gran número de peruanos salieron a las calles en contra del golpe de Estado

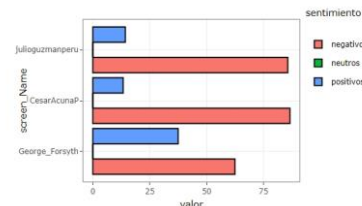
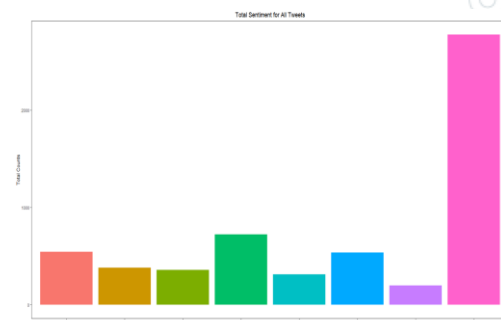
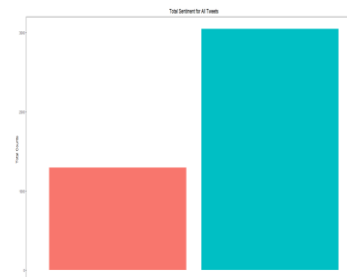
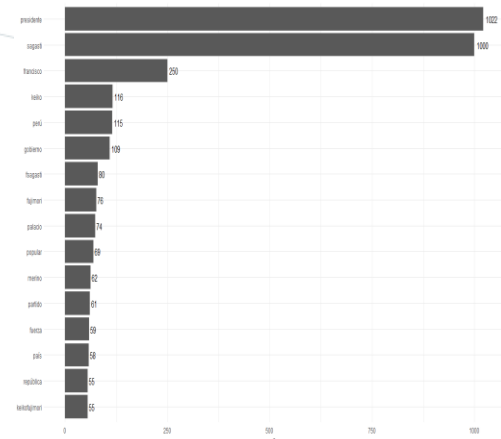
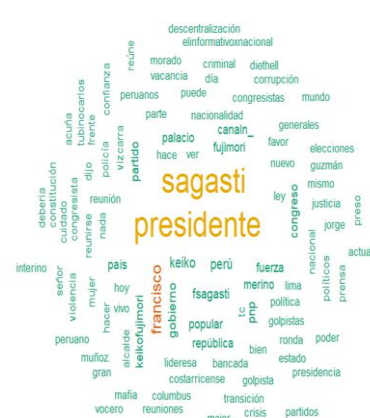
### Fallecimiento de Inti Sotelo y Bryan Pintado

**Francisco Sagasti juró el 17 de noviembre como presidente del Perú**

**Manuel Merino**  
asume un  
Gobierno  
de facto

## Pese a las manifestaciones, Merino de Lama nombró un gabinete ministerial

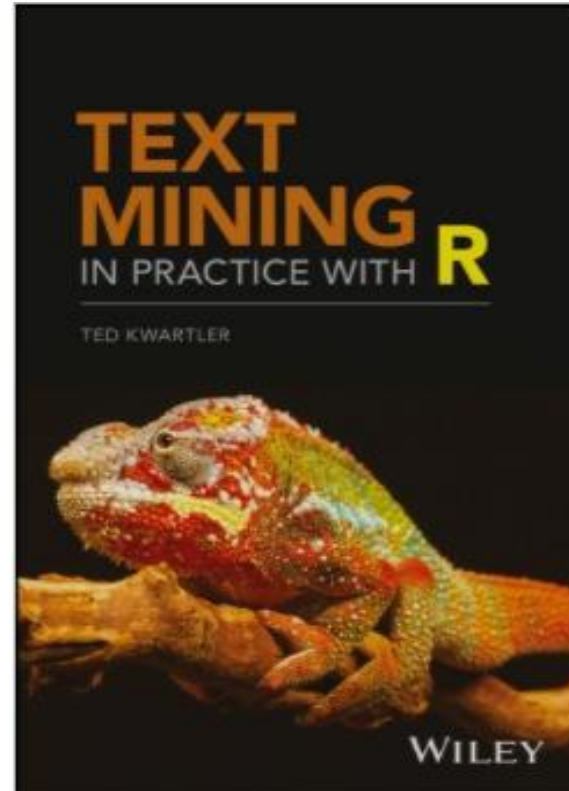
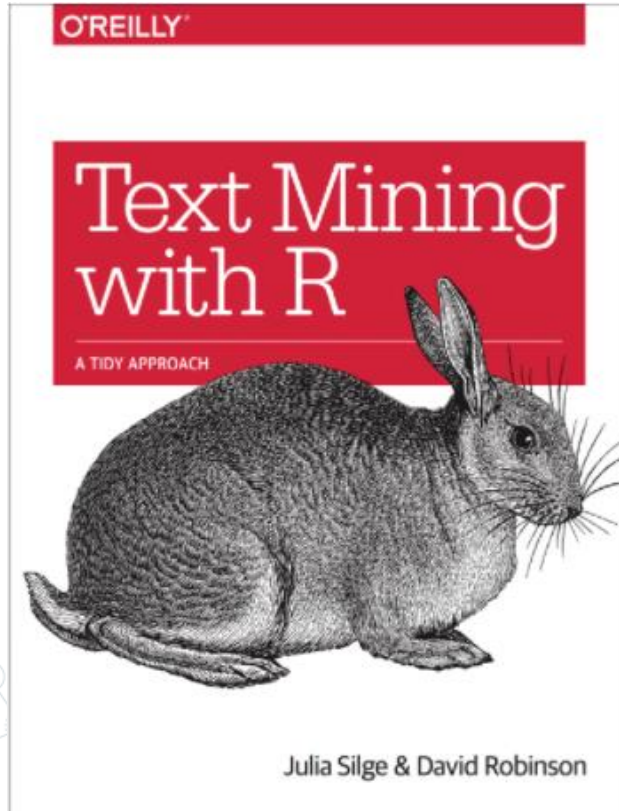
## Renuncia de Merino



Data Ladies Perú



## Libros !





# Thanks!

## Any questions?

**Estephani** Rivera Jaramillo

CONSULTOR ESTADÍSTICO, ANALISTA COMERCIAL

estephani.rivera.j@gmail.com 

924440966 

San Juan de Miraflores, Lima 

Estephani Rivera Jaramillo 