

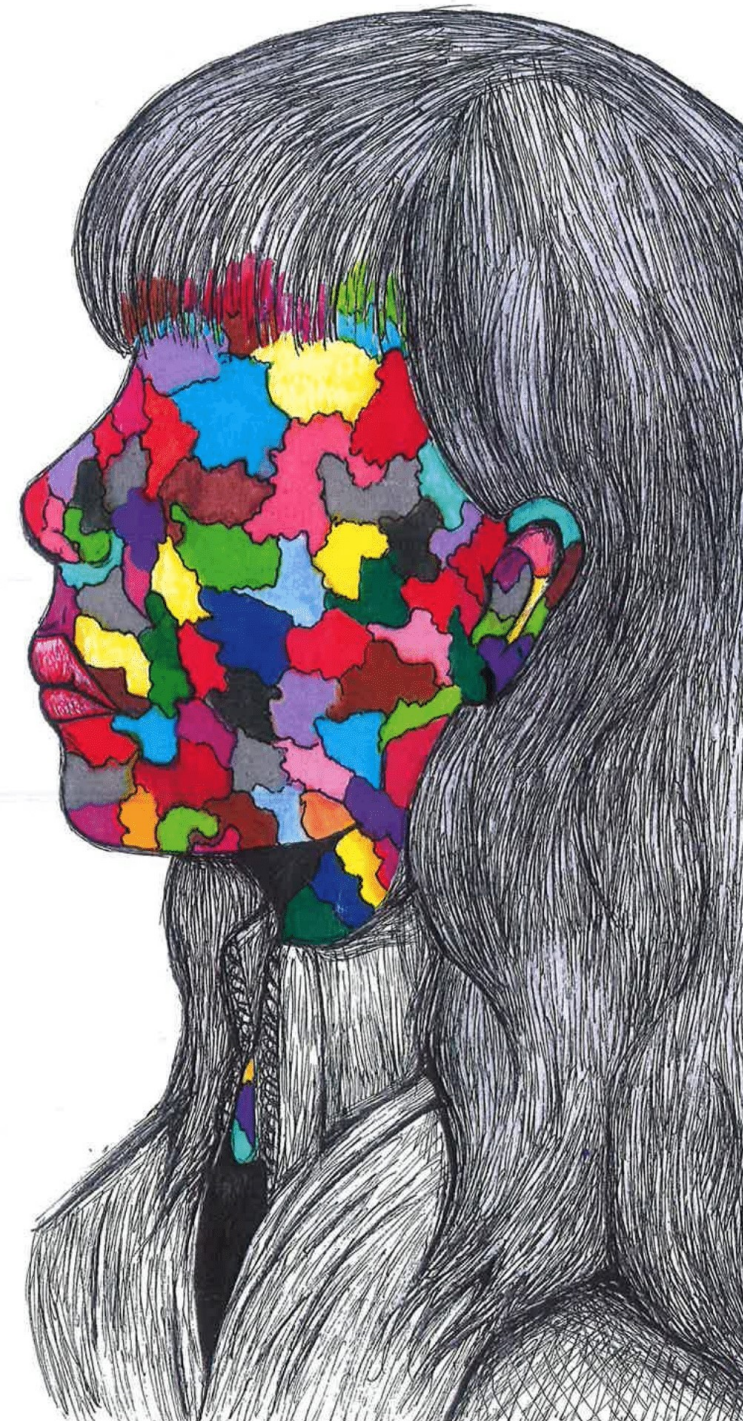


Máster de Sociología Aplicada a la
Investigación Social y de Mercados
Universidad de A Coruña
Convocatoria: Septiembre, 2019

**ENFOQUE INTERDISCIPLINAR SOBRE
EL ESTUDIO CUANTITATIVO DE LAS
EMOCIONES**

Anastasiya Shevchenko

Directores:
María del Carmen Rodríguez Rodríguez
Rubén Fernández Casal



INTRODUCCIÓN

Objeto de estudio: El análisis de las emociones con técnicas de procesamiento automático de texto, conocido como **.sentiment analysis** y aplicable a texto escrito en castellano

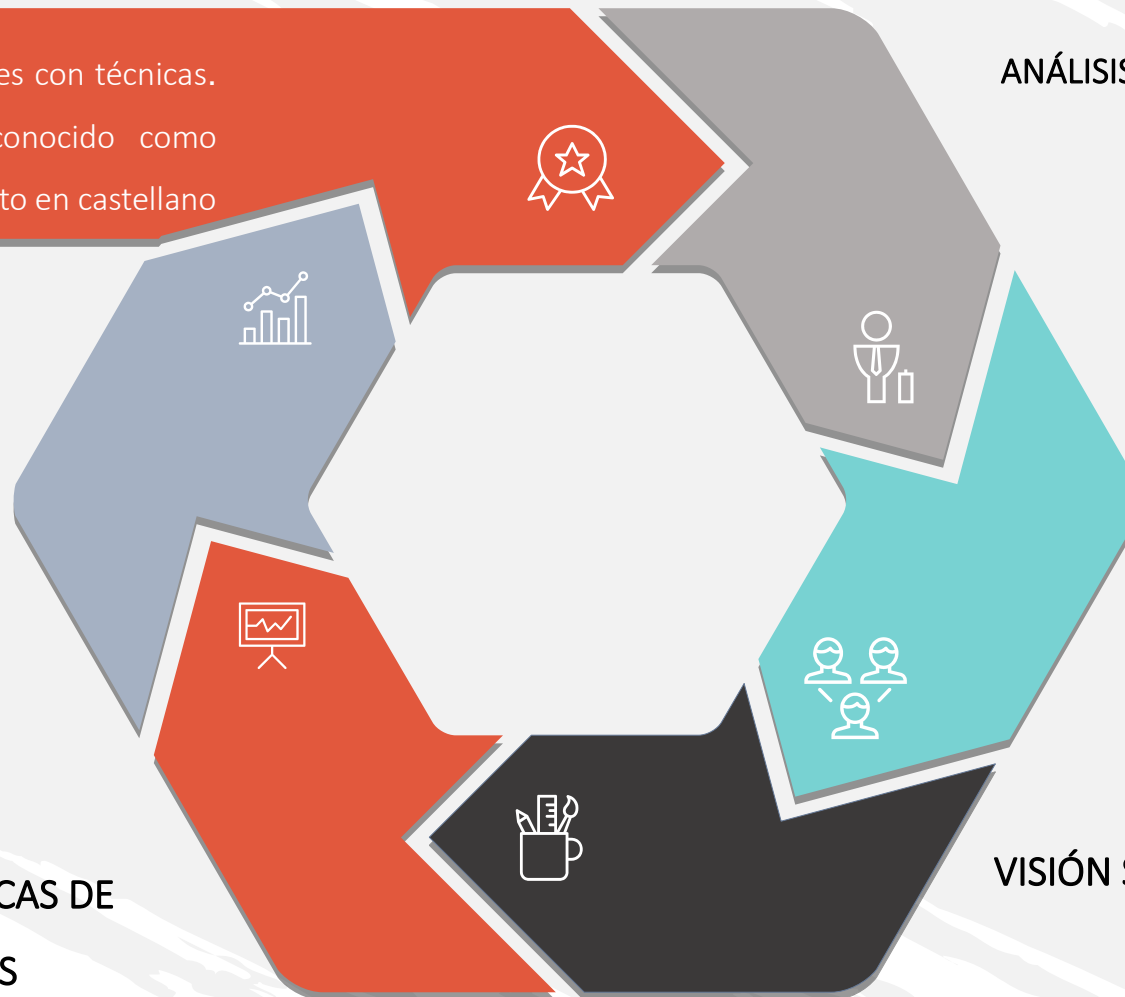
ANÁLISIS CUANTITATIVO DE LAS EMOCIONES
EN LA EMPRESA

NUEVAS METODOLOGÍAS PARA
SOCIOLOGÍA

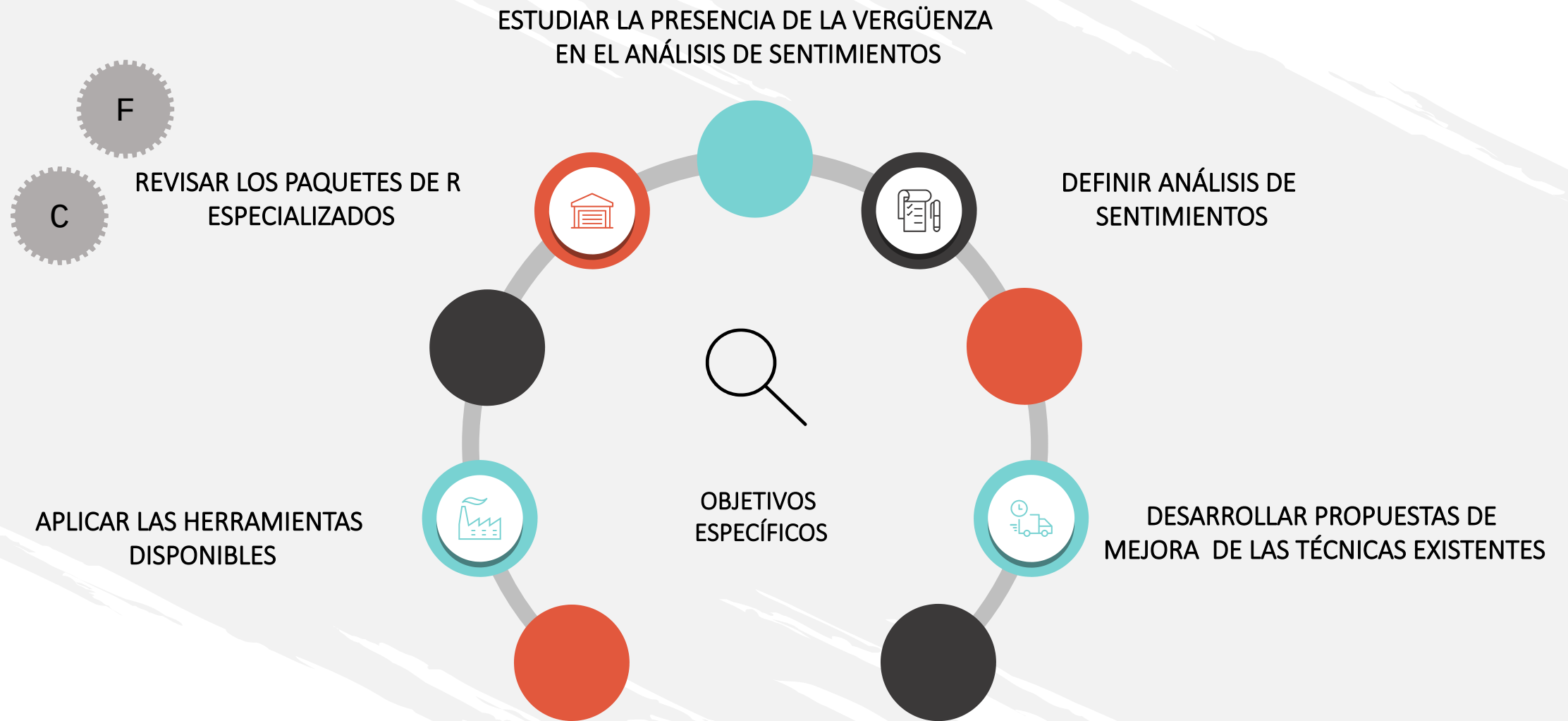
DESCONEXIÓN ENTRE CIENCIAS

ESTADO Y UTILIZACIÓN DE TÉCNICAS DE
ANÁLISIS DE SENTIMIENTOS

VISIÓN SOCIOLÓGICA EN LA CONSTRUCCIÓN
DE TÉCNICAS



OBJETIVOS



Revisar las técnicas de análisis de sentimientos aplicables a castellano, aplicándolas a un conjunto de datos real y desarrollando propuestas para su mejora..

ANÁLISIS DE SENTIMIENTOS

POLARITY

Conocimiento

DATOS NO SUPERVISADOS

SEMÁNTICO

SINTÁCTICO

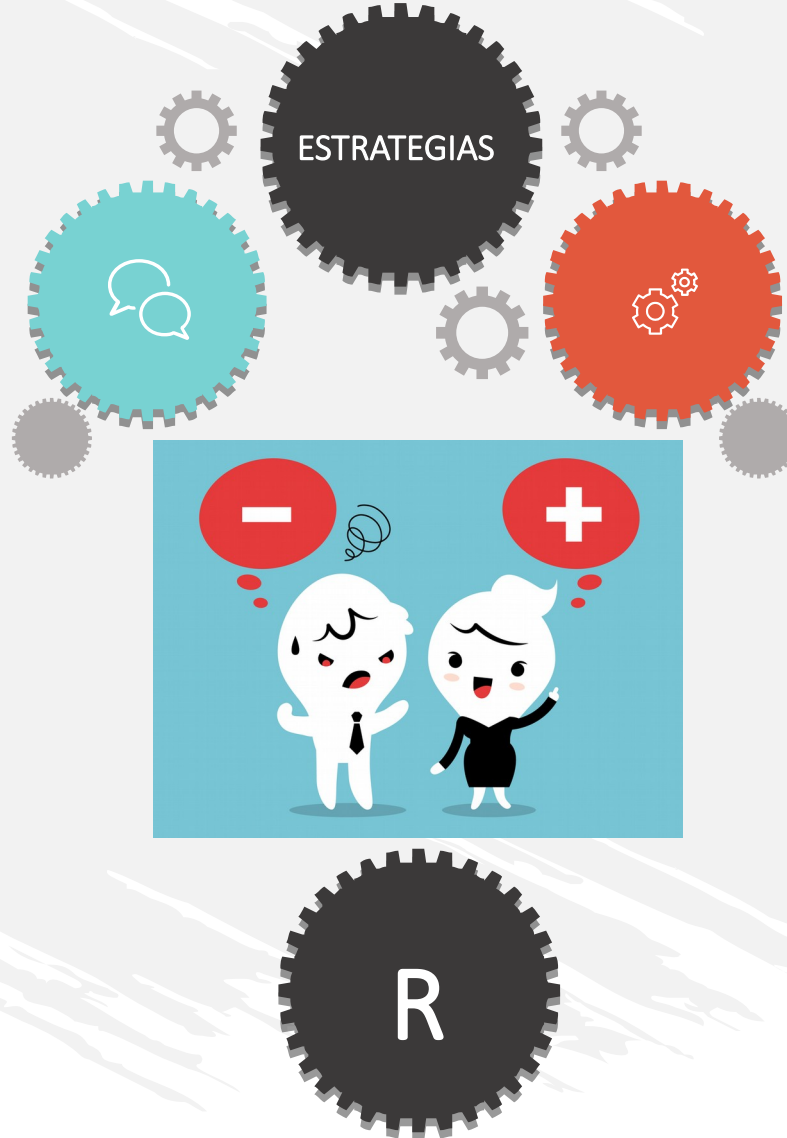
ESTRATEGIAS

Aprendizaje automático

DATOS SUPERVISADOS

ALGORITMOS

R



JUSTIFICACIÓN

APLICACIONES

Análisis de reviews

Estudios de inteligencia

Sistemas de recomendación



Empresas



ACTUALMENTE EL CENTRO DEL
ANÁLISIS **NO** SON LAS EMOCIONES

UTILIDADES

Gestión de grandes
volumenes de datos

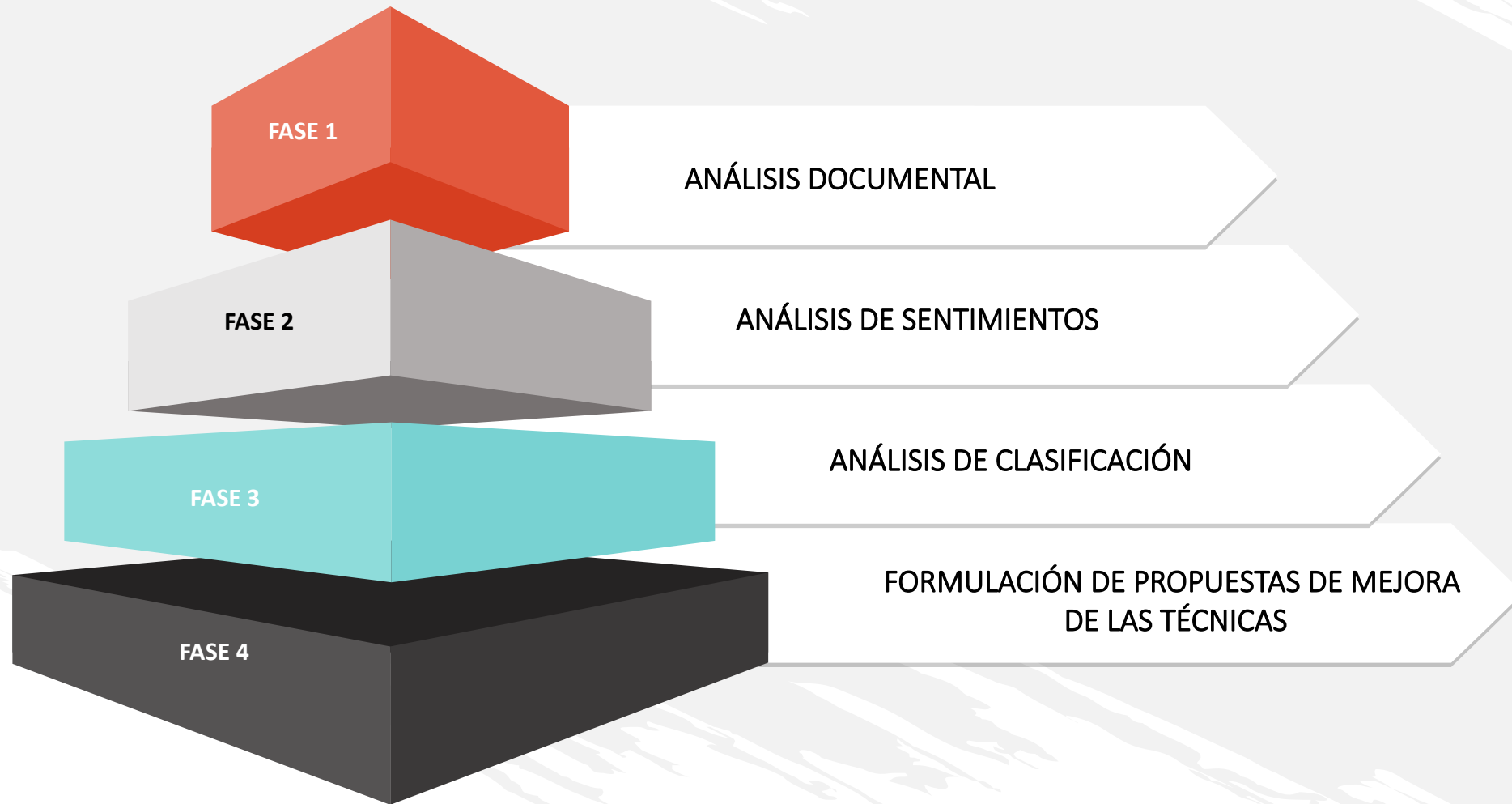
Extracción de información
relevante

Reducción de procesamiento
manual



Sociología

METODOLOGÍA

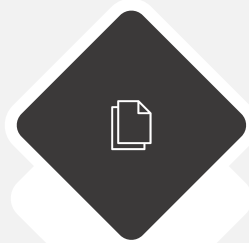


FASES: ANÁLISIS DE SENTIMIENTOS

PASO 01
MANIPULACIÓN DE TEXTO
Organización de los datos



PASO 03
MATRIZ DE DOCUMENTOS TÉRMINO
Separación por términos



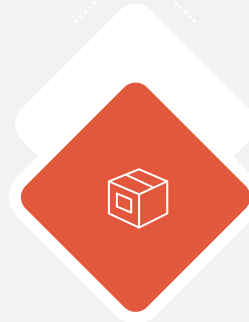
PASO 05
APLICACIÓN DE POLARITY
Cálculo de polaridad



PASO 02
PREPROCESAMIENTO
Depuración y tipificación

PASO 04
EXPLORACIÓN
Frecuencias, nubes, correlaciones

PASO 06
EXTRACCIÓN
Organización de puntuaciones





INTENSIFICADORAS

DISMINUIDORAS

NEGADORES

POLARITY

iSol

+ 2509;- 5626

Considerando que la palabra positiva se puntúa como +1 y la palabra negativa como -1, se calculan los siguientes valores, donde W es el peso de la palabra polarizada y c es el peso de los intensificadores.

$$W_i^{neg} = N_i \bmod 2$$

$$Dism = \max(-c \cdot (D_i + W_i^{neg} \cdot A_i), -1)$$

$$Amp = c \cdot A_i \cdot (1 - W_i^{neg})$$

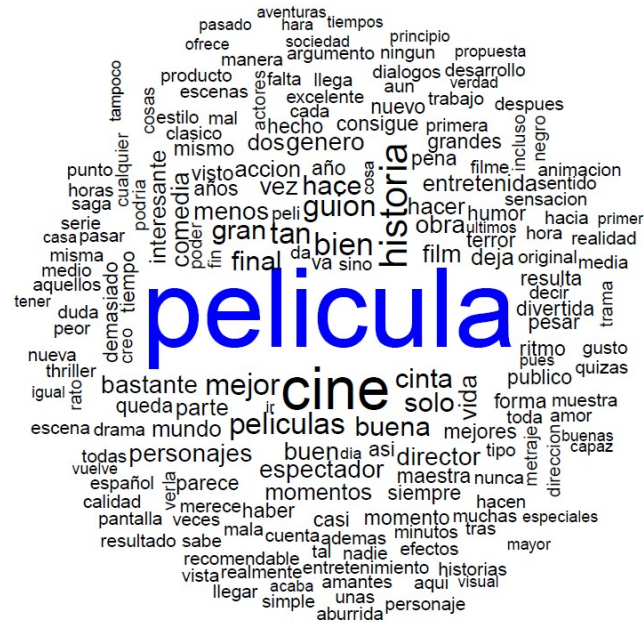
El total de la polaridad se calcula para un clúster, formado a partir de una cantidad definida de palabras alrededor de cada palabra que contiene una carga positiva o negativa.

$$C_i = (1 + (Dism + Amp)) \cdot W_i \cdot (-1)^{N_i}$$

Por último, para obtener una única puntuación por cada comentario o dato de texto, se lleva a cabo una suma de las puntuaciones de los clústers y se divide por la raíz cuadrada del total de las palabras dentro de la cadena de texto.

$$C = \frac{\sum C_i}{\sqrt{n}}$$

Frecuencia de palabras con representación gráfica



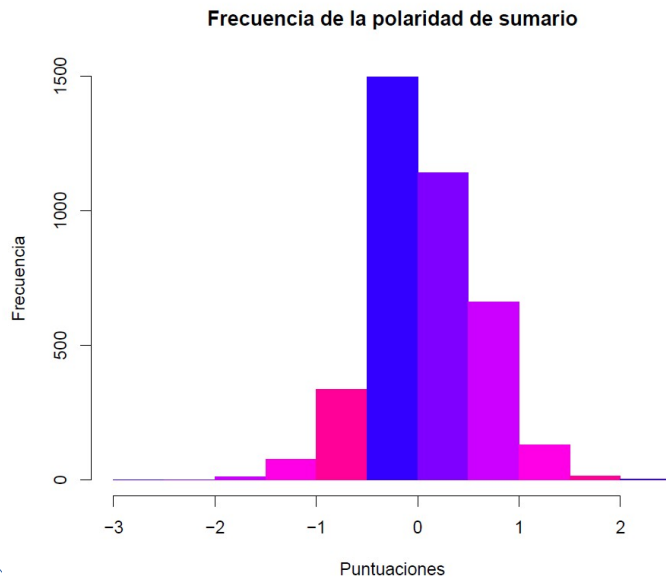
Las 15 correlaciones más fuertes de 'sumario'

```
mcor_sum_desc[1:10,]
```

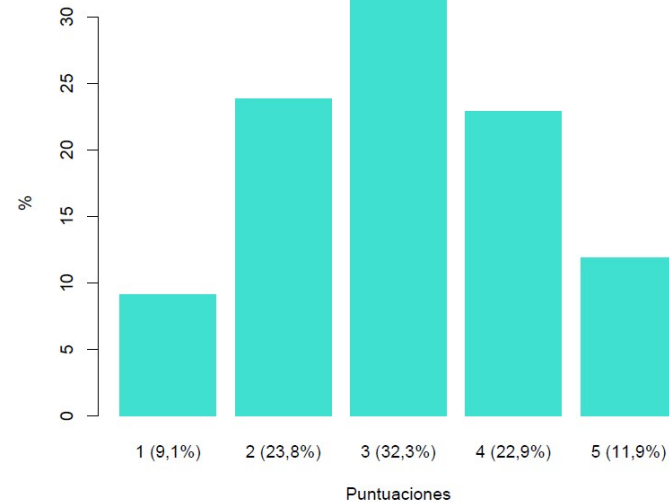
##	row	column	cor	p
## 1	especiales	efectos	0.8421792	0
## 2	efectos	especiales	0.8421792	0
## 3	obra	maestra	0.7576595	0
## 4	maestra	obra	0.7576595	0
## 5	rato	pasar	0.5891669	0
## 6	pasar	rato	0.5891669	0
## 7	pena	merece	0.5583780	0
## 8	merece	pena	0.5583780	0

RESULTADOS DE POLARIDAD

Histograma de puntuaciones de polaridad calculadas



Puntuaciones de valoración originales



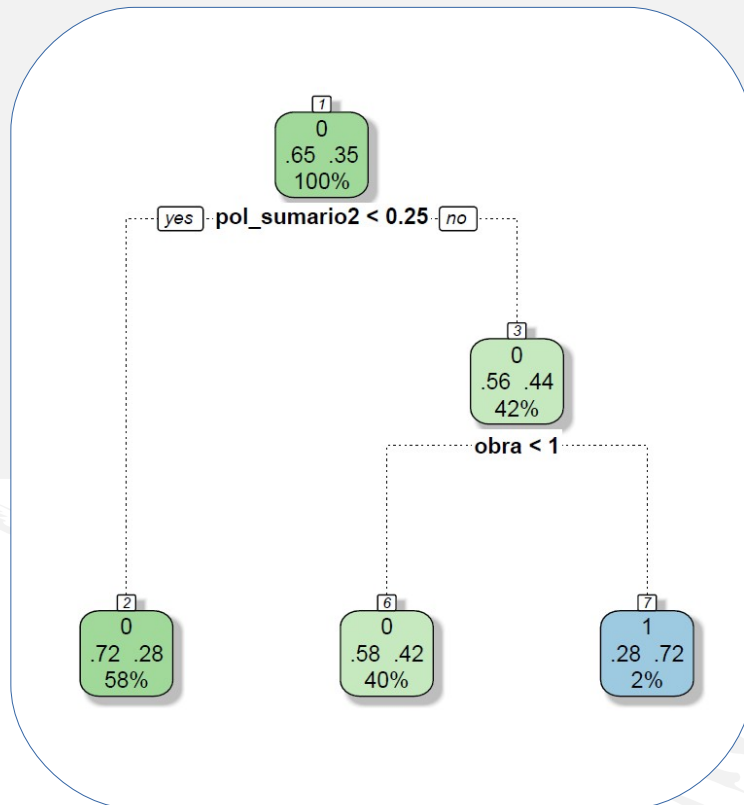
Salida obtenida a partir de la función

```
str(res_polaridad_texto[[1]], width = 70, strict.width = 'wrap')

## 'data.frame': 1 obs. of 6 variables:
## $ all : chr "all"
## $ wc : int 266
## $ polarity : num 0.429
## $ pos.words:List of 1
## ..$ : chr "amigo" "pacifica" "amigo" "prima" ...
## $ neg.words:List of 1
## ..$ : chr "terror" "cara" "terror" "locura" ...
## $ text.var : chr "may quieres amigo peliculas recuerdan terror
## siempre lleva garras acero mano mascara cara terror locura
## encuent"| __truncated__
```

RESULTADOS DE CLASIFICACIÓN

Árbol de decisión



Bosque aleatorio

```
## Type of random forest: classification
## Number of trees: 500
## No. of variables tried at each split: 13
##
## OOB estimate of error rate: 34.05%
## Confusion matrix:
##      0   1 class.error
## 0 1452 313  0.1773371
## 1  611 338  0.6438356

##           0   1 MeanDecreaseAccuracy MeanDecreaseGini
## pol_sumario2 28.25  8.61                28.56          61.73
## obra         20.77 13.97                25.74           8.71
## buena        17.18 -0.78                14.44           5.31
## mejores      13.79  3.44                13.41           5.36
## hacia         8.20  9.47                11.58           3.02
## año           6.75  9.34                11.32           4.66
```

CONCLUSIONES



PROPUESTAS DE MEJORA



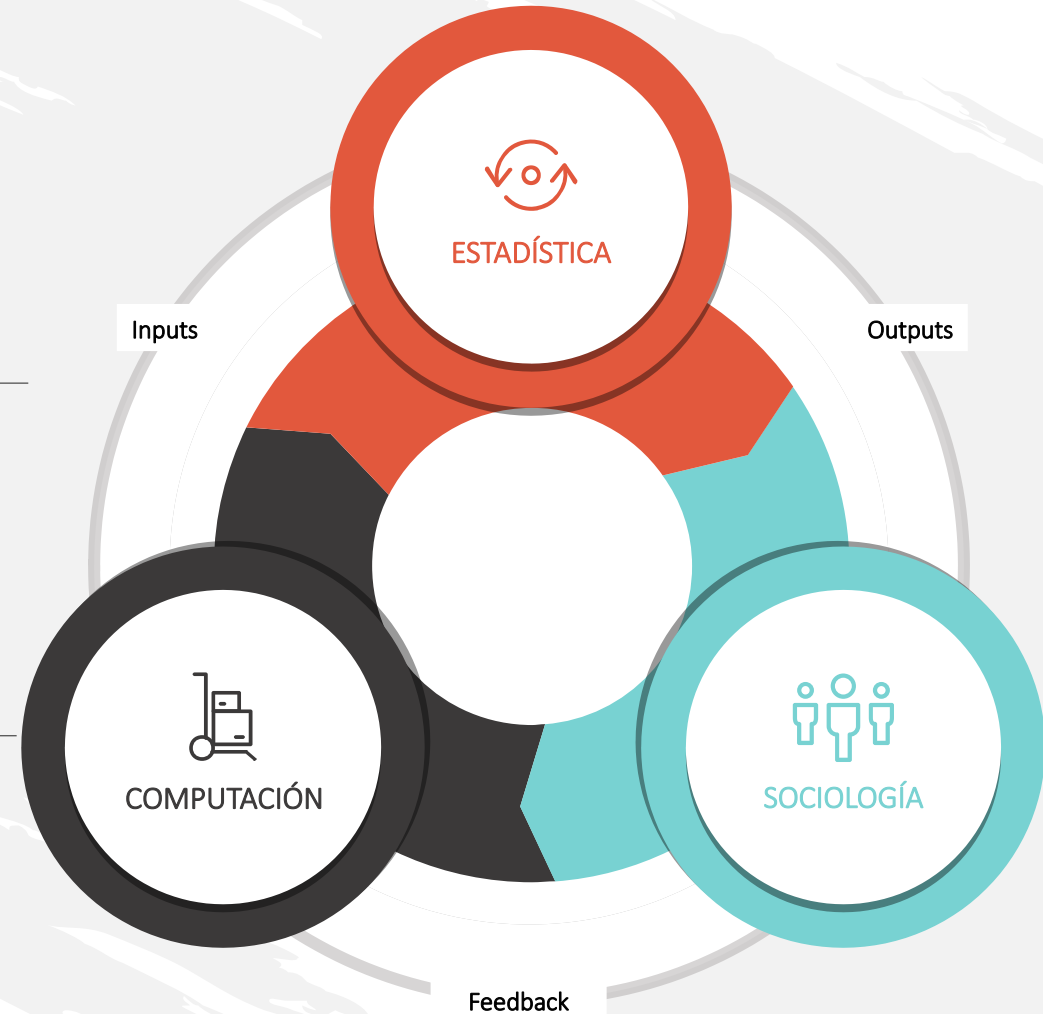
1. MEJORA DE LA DEPURACIÓN DE LOS DATOS.
2. IMPLEMENTACIÓN DE NUEVOS MÉTODOS DE OBTENCIÓN DE RESULTADOS.
3. AUTOMATIZACIÓN DE PROCESOS.
4. EXTRACCIÓN DE DICCIONARIOS TEMÁTICOS.
5. CREACIÓN DE INTERFAZ Y RECURSOS DE SOPORTE: *SHINY*.



1. DESARROLLO DE NUEVOS ESTIMADORES.
2. PONDERACIÓN DE LAS EMOCIONES, ASÍ COMO SU INTENSIDAD.
3. COMBINACIÓN DE LAS UNIDADES DE MEDIDA DE LOS RECURSOS LÉXICOS.



1. TEORIZACIÓN DE LAS EMOCIONES, EVITANDO LA POLARIZACIÓN.
2. ELABORACIÓN DE DICCIONARIOS EN CASTELLANO SOBRE LAS EMOCIONES.
3. DIMENSIONALIZACIÓN DE LAS EMOCIONES.
4. APROXIMACIÓN Y USO DEL ANÁLISIS DE SENTIMIENTOS.
5. SUPERVISIÓN DE CONJUNTOS DE DATOS.



CONSIDERACIONES FINALES





"The best and most beautiful things in the world cannot be seen or even touched. They must be felt with the heart"

Helen Keller

