



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación



El futuro
es de todos

DNP
Departamento
Nacional de Planeación

Identificación de percepción de competencia desleal en sectores de la economía colombiana

Carlos Andrés Rocha Ruiz - UCD
Dirección de Desarrollo Digital

febrero, 2021



1. ¿Qué es la UCD?
2. ¿Qué proyectos hemos trabajado?
3. Analítica aplicada a políticas públicas
4. Competencia desleal
5. Fuentes de información
6. Metodología
7. Resultados



1. ¿Qué es la UCD?



¿Quiénes somos?

Somos el primer equipo del sector público dedicado a la **explotación de datos** de distintas fuentes y en grandes volúmenes para la formulación de **políticas públicas** en Colombia



5

¿Qué hacemos?

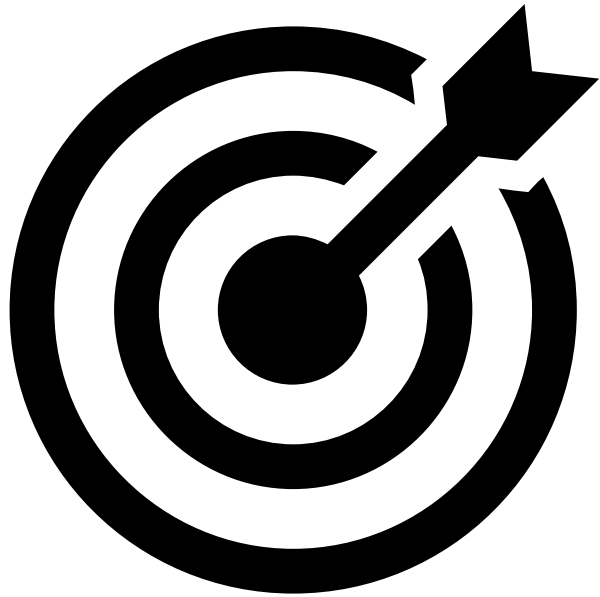
Utilizamos **analítica de datos** para brindar insumos que orienten la toma de decisiones en el marco de la formulación de políticas públicas



El futuro
es de todos

DNP
Departamento
Nacional de Planeación

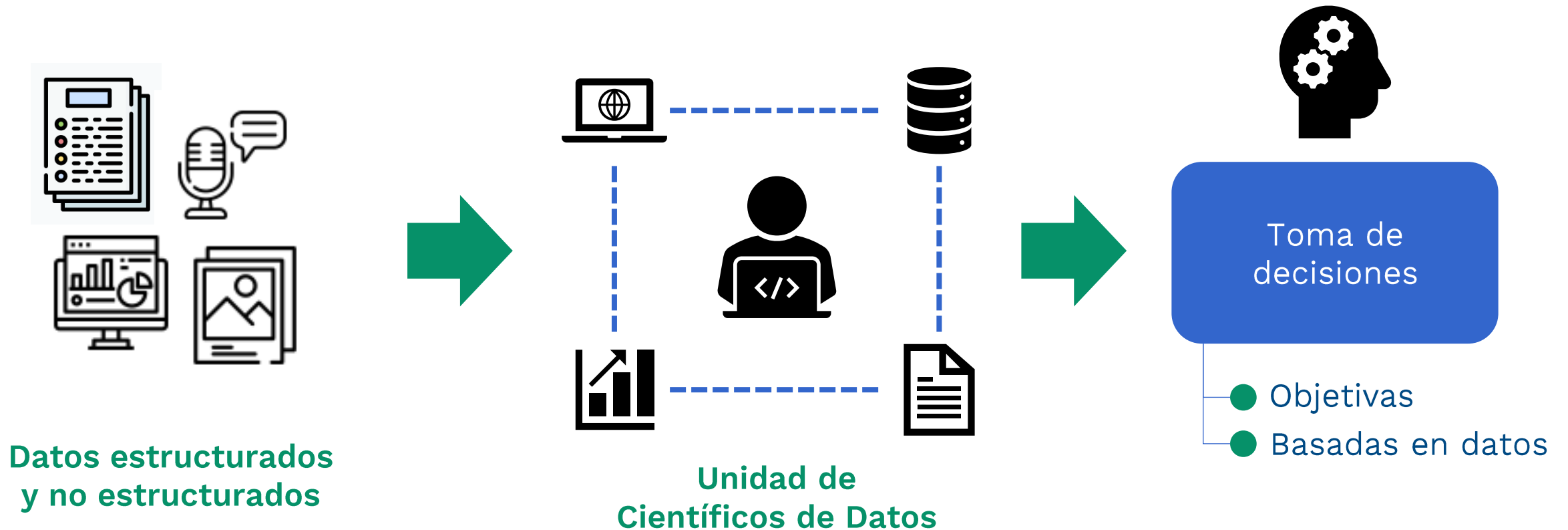
Misión



Fortalecer la creación de valor en el sector público mediante el desarrollo de proyectos de Big Data y analítica de datos que permitan el aprovechamiento efectivo de datos estructurados y no estructurados para la **toma de decisiones** objetiva.



Vista general de la UCD



Tipos de datos que utilizamos

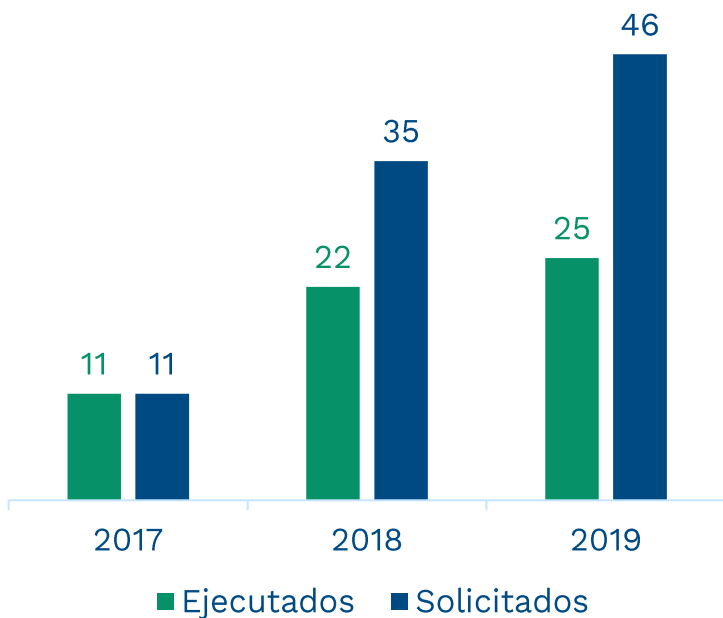


2. ¿Qué proyectos hemos trabajado?



Algunos proyectos

Proyectos y pilotos de analítica de datos



Externos



Detección de vías terciarias con imágenes satelitales (MinTransporte)



Análisis de prescripciones sobre la base de MIPRES no PBS (ADRES)

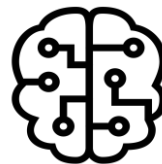


Identificación de patrones de reincidencia y reiteración delictiva (Fiscalía)



Aplicativo para el análisis y caracterización automática de PQRS (UARIV)

Internos



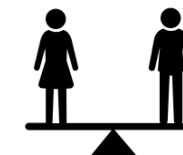
Inteligencia artificial para priorizar seguimiento a proyectos de Regalías.



Identificación de asentamientos en riesgo de inundación



Identificación de inversión pública para el cumplimiento de los ODS



Estimación de la brecha salarial de género en los contratos públicos.



El futuro es de todos

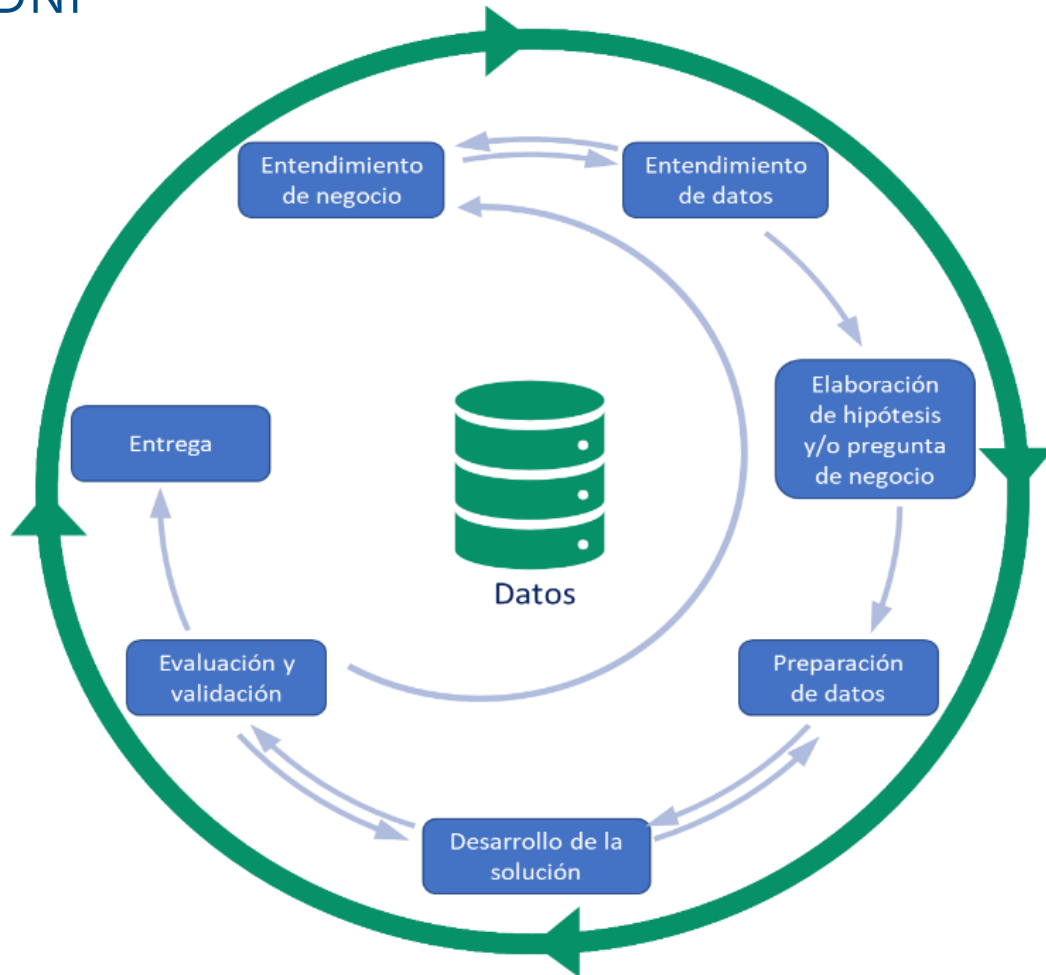
DNP
Departamento
Nacional de Planeación

3. Analítica aplicada a políticas públicas



Políticas públicas y analítica

DNP



Identificación del problema

Identificación de las fuentes de información

Planteamiento del proyecto

Desarrollo del proyecto

Despliegue

Políticas públicas y analítica

DNP



Entendimiento
del negocio

Identificación del problema

Identificación de las
fuentes de información

Planteamiento del proyecto

Desarrollo del proyecto

Despliegue

Políticas públicas y analítica

DNP

Entendimiento
de los datos

Identificación del problema

Identificación de las
fuentes de información

Planteamiento del proyecto

Desarrollo del proyecto

Despliegue

Políticas públicas y analítica

DNP

Elaboración de
hipótesis y/o
preguntas de
negocios

Identificación del problema

Identificación de las
fuentes de información

Planteamiento del proyecto

Desarrollo del proyecto

Despliegue

Políticas públicas y analítica

DNP



Identificación del problema

Identificación de las fuentes de información

Planteamiento del proyecto

Desarrollo del proyecto

Despliegue

Políticas públicas y analítica

DNP



Identificación del problema

Identificación de las
fuentes de información

Planteamiento del proyecto

Desarrollo del proyecto

Despliegue

4. Competencia desleal



¿Por qué es importante?



Dirección de Innovación y Desarrollo
Empresarial

5. Fuentes de información



Descarga de información

Twitter y portales de noticias



Descarga de información

Twitter y portales de noticias

API



Web Scraping



Descarga de información

Twitter y portales de noticias

API



Web Scraping



Descarga de información

Twitter y portales de noticias

API



Web Scraping



¿Por qué Selenium y no beautifulsoup?

Descarga de información

Twitter y portales de noticias



Automatización y
actualización periódica

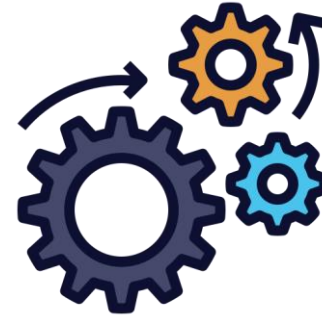
6. Metodología



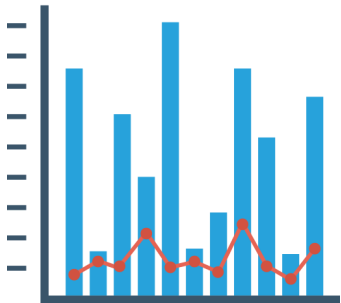
Metodología del proyecto



Fase 1
Proceso de descarga
de tweets por
término



Fase 2
Procesamiento de texto
(limpieza de textos,
eliminación de
stopwords y términos
claves)



Fase 3
Primera
visualizaciones
(nube de palabras,
conurrencia,
frecuencias)

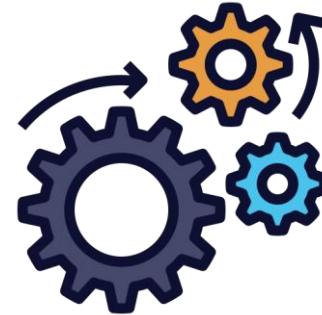


Fase 4
Visualización de los
resultados en un
reporte

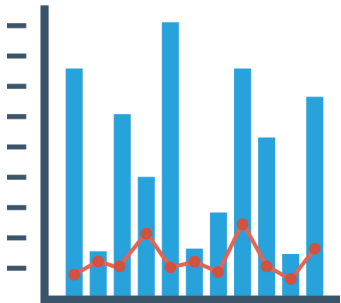
Metodología del proyecto



Fase 1
Proceso de descarga
de tweets por
términos



Fase 2
Procesamiento de texto
(limpieza de textos,
eliminación de
stopwords y términos
claves)



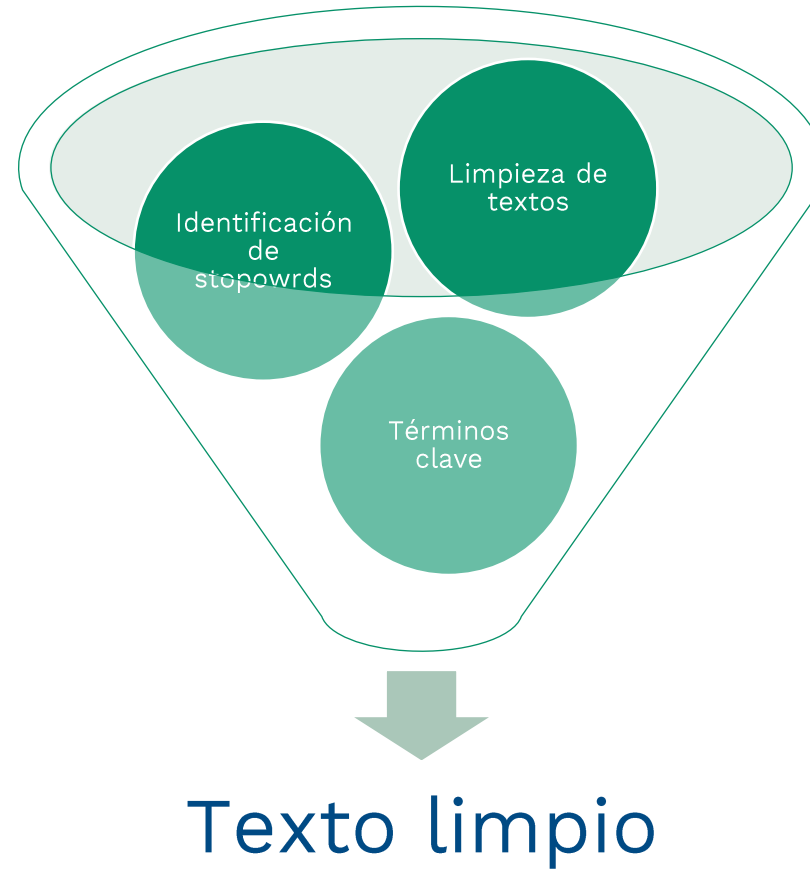
Fase 3
Primera
visualizaciones
(nube de palabras,
conurrencia,
frecuencias)



Fase 4
Visualización de los
resultados en un
reporte

Metodología del proyecto

Fase 2

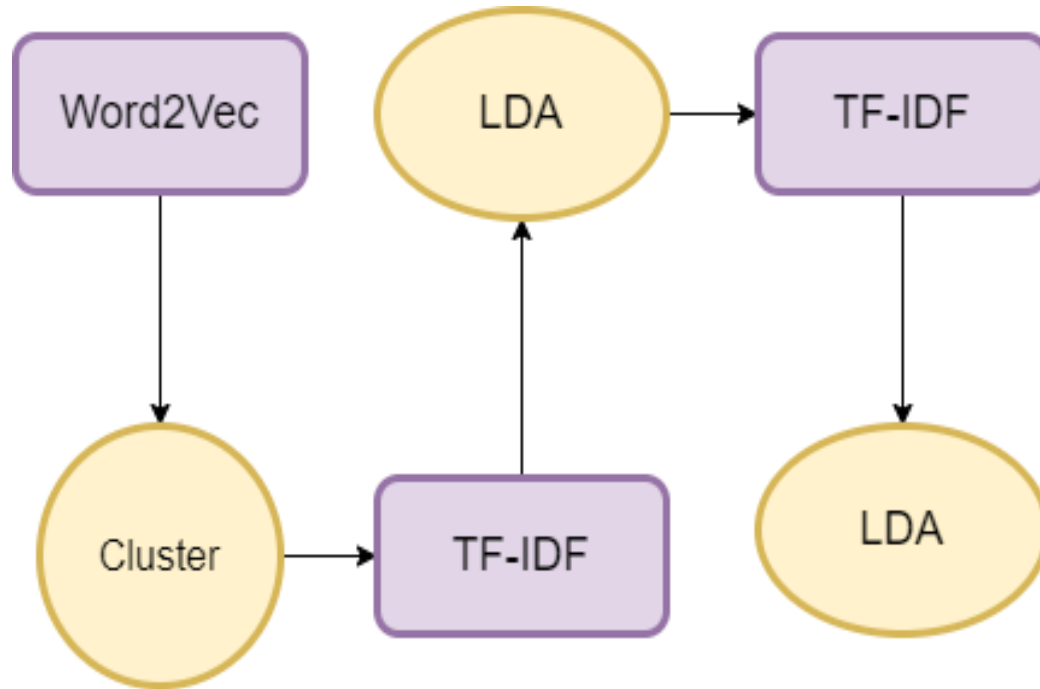


Metodología del proyecto

Fase 2

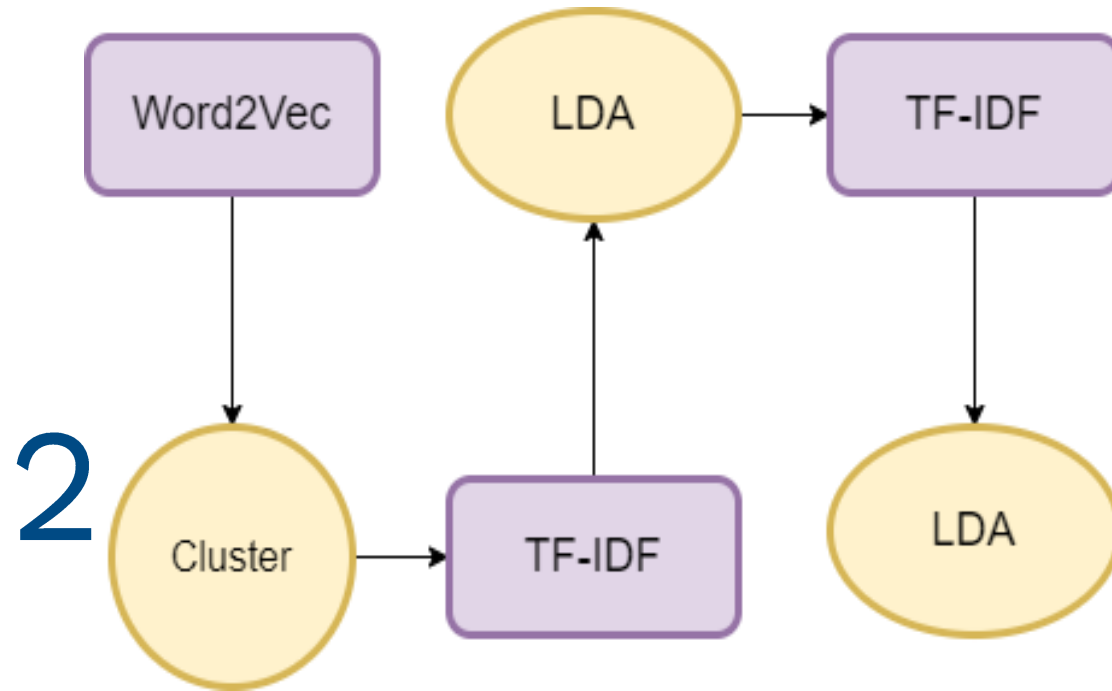


1



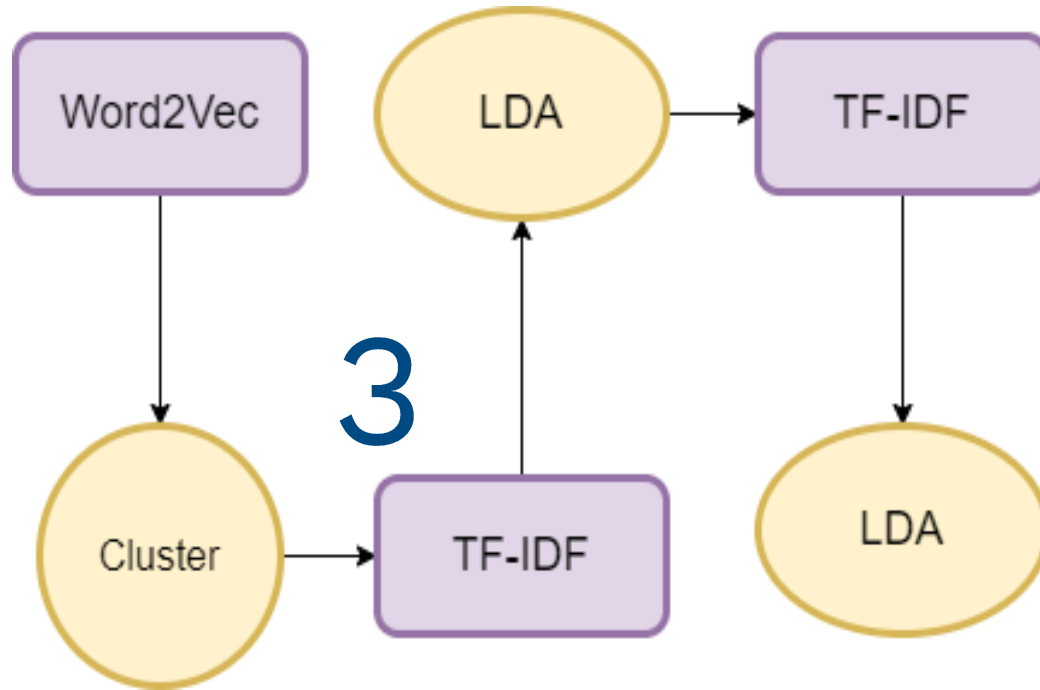
Metodología del proyecto

Fase 2



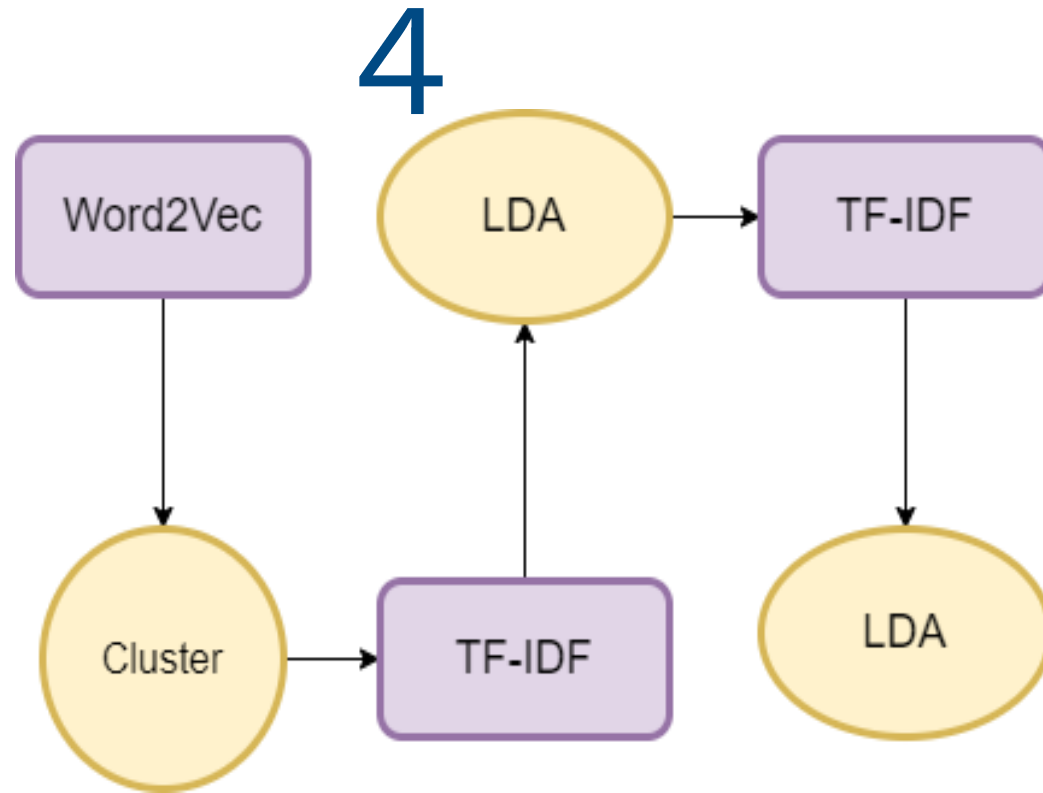
Metodología del proyecto

Fase 2



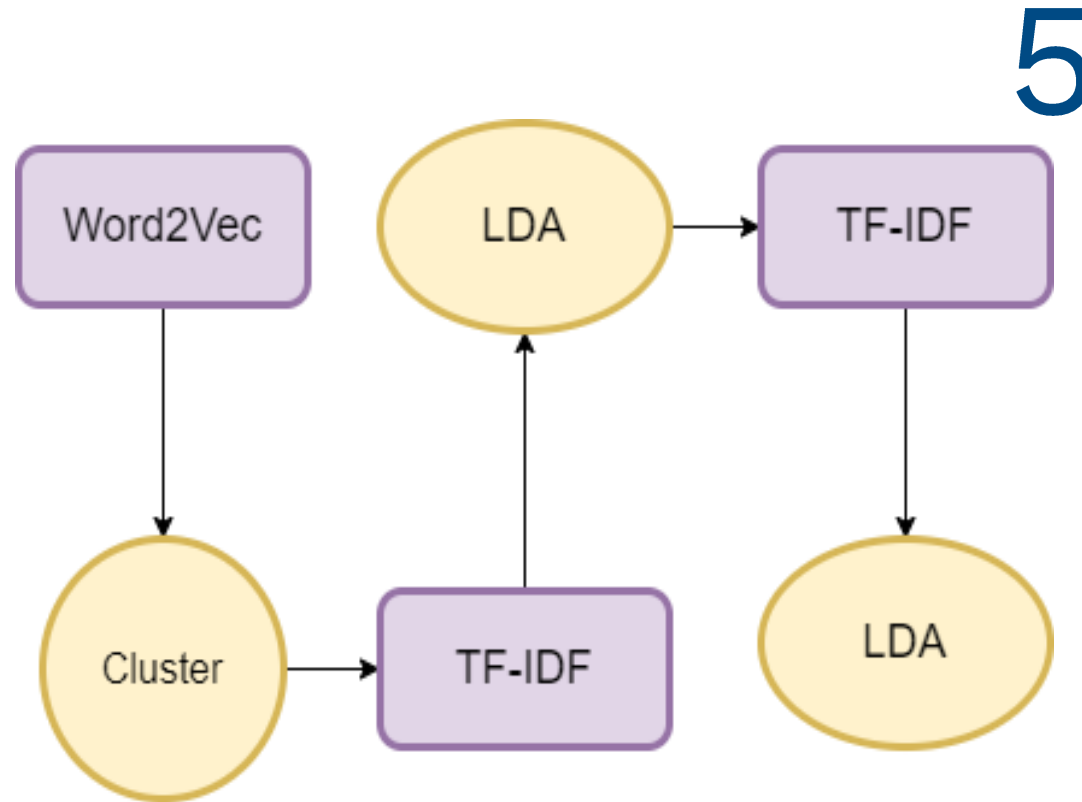
Metodología del proyecto

Fase 2



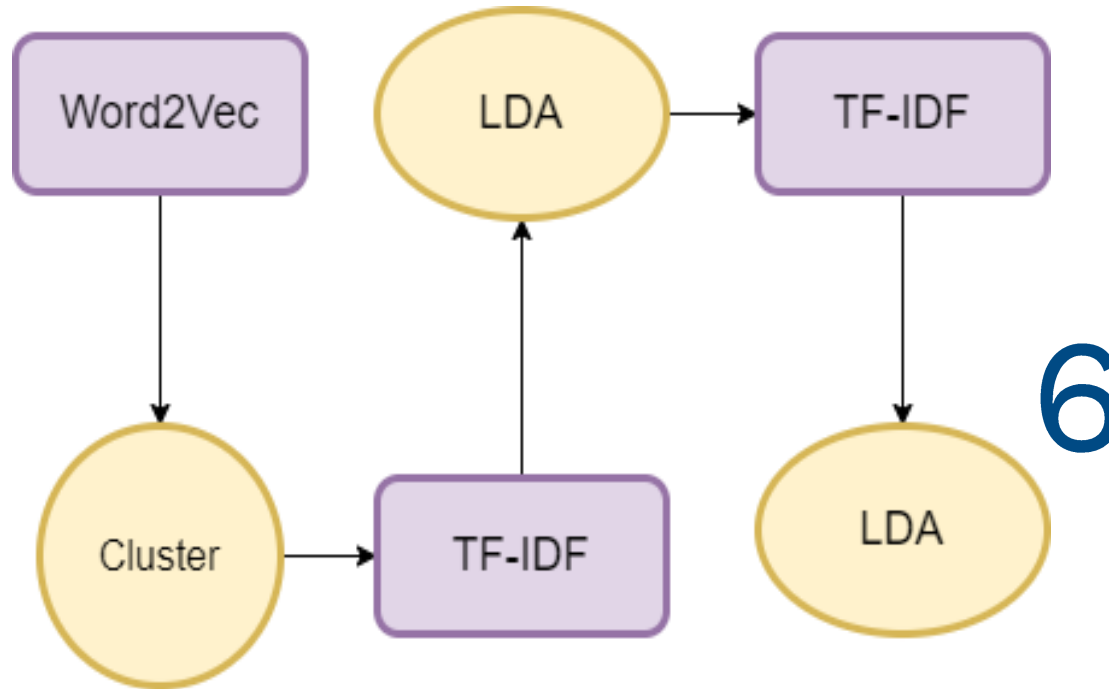
Metodología del proyecto

Fase 2



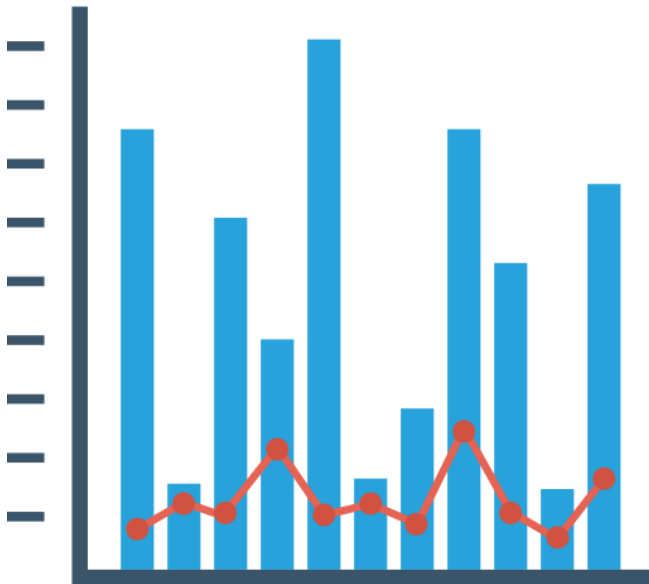
Metodología del proyecto

Fase 2



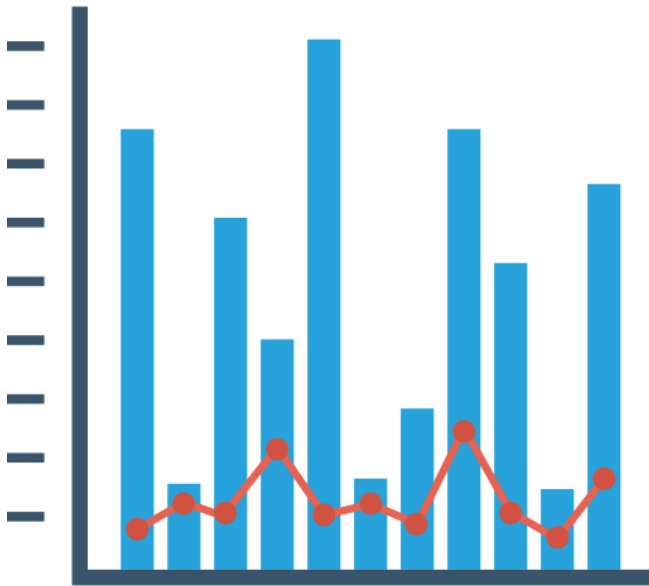
Visualizaciones

Entre clustering y resultados



Visualizaciones

Entre clustering y resultados



Nubes de palabras

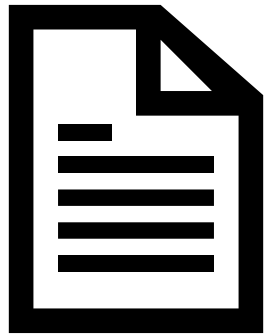


Aplicativo

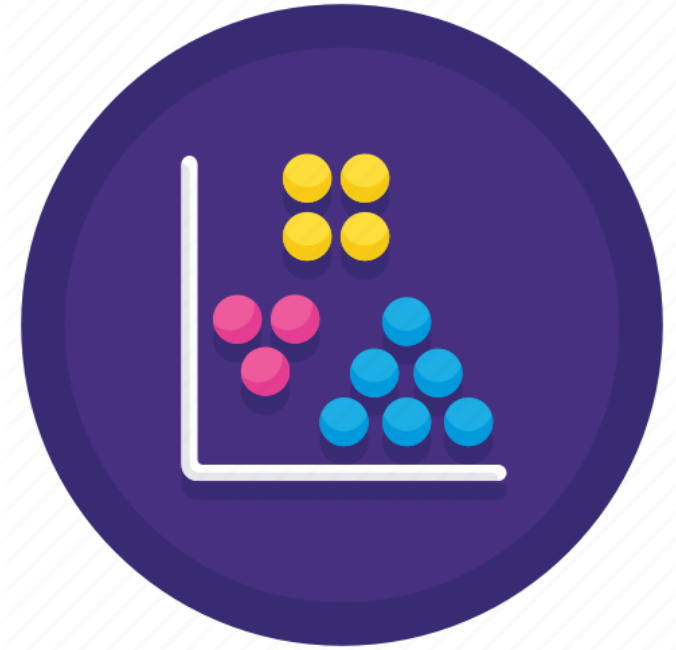
7. Resultados



Identificación de sectores

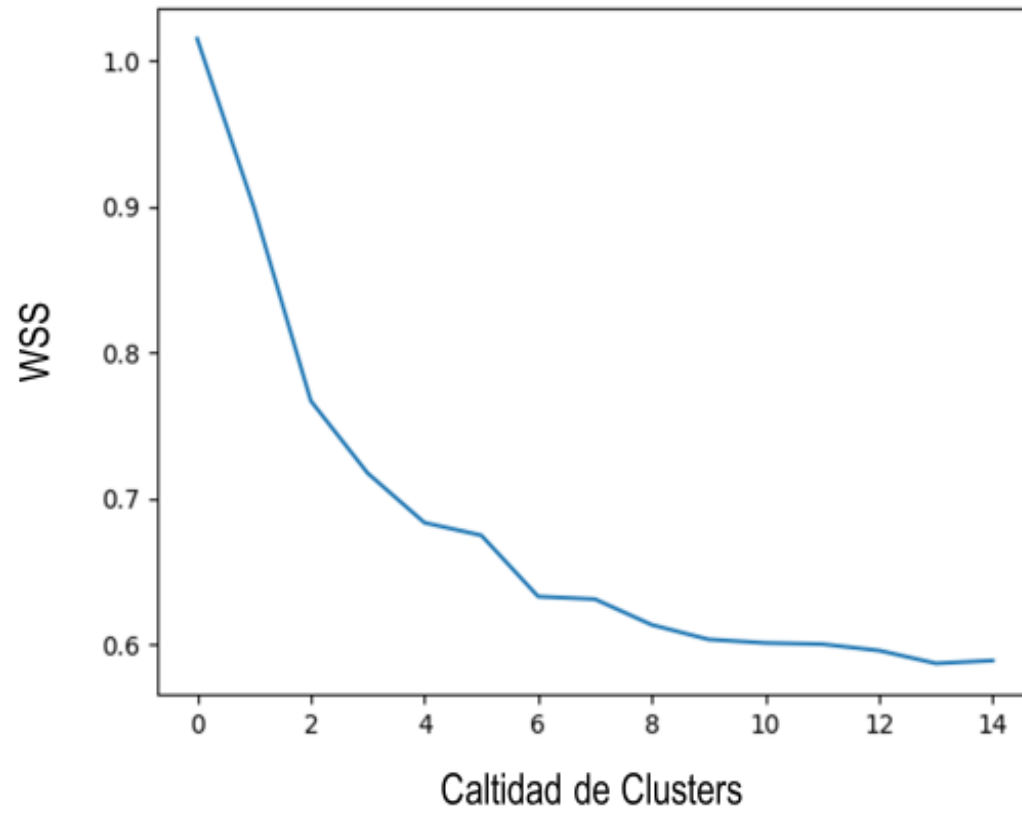


Word2Vec

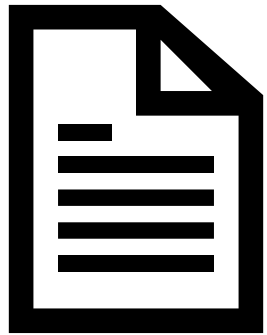


Medida WSS

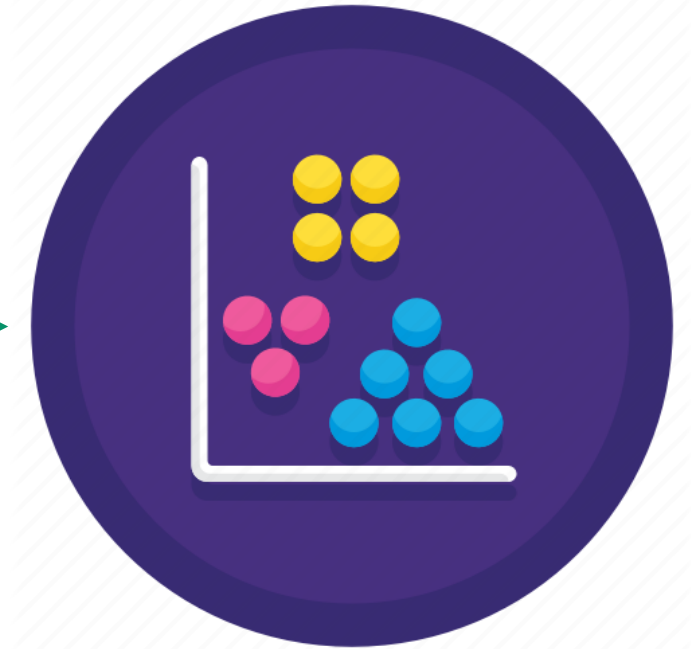
Primer Cluster



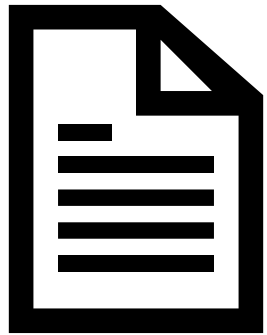
Identificación de sectores



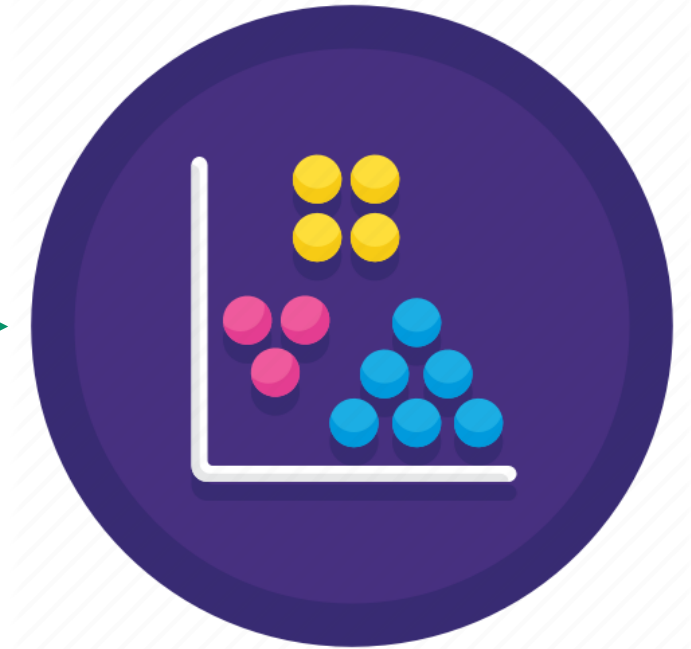
TF-IDF



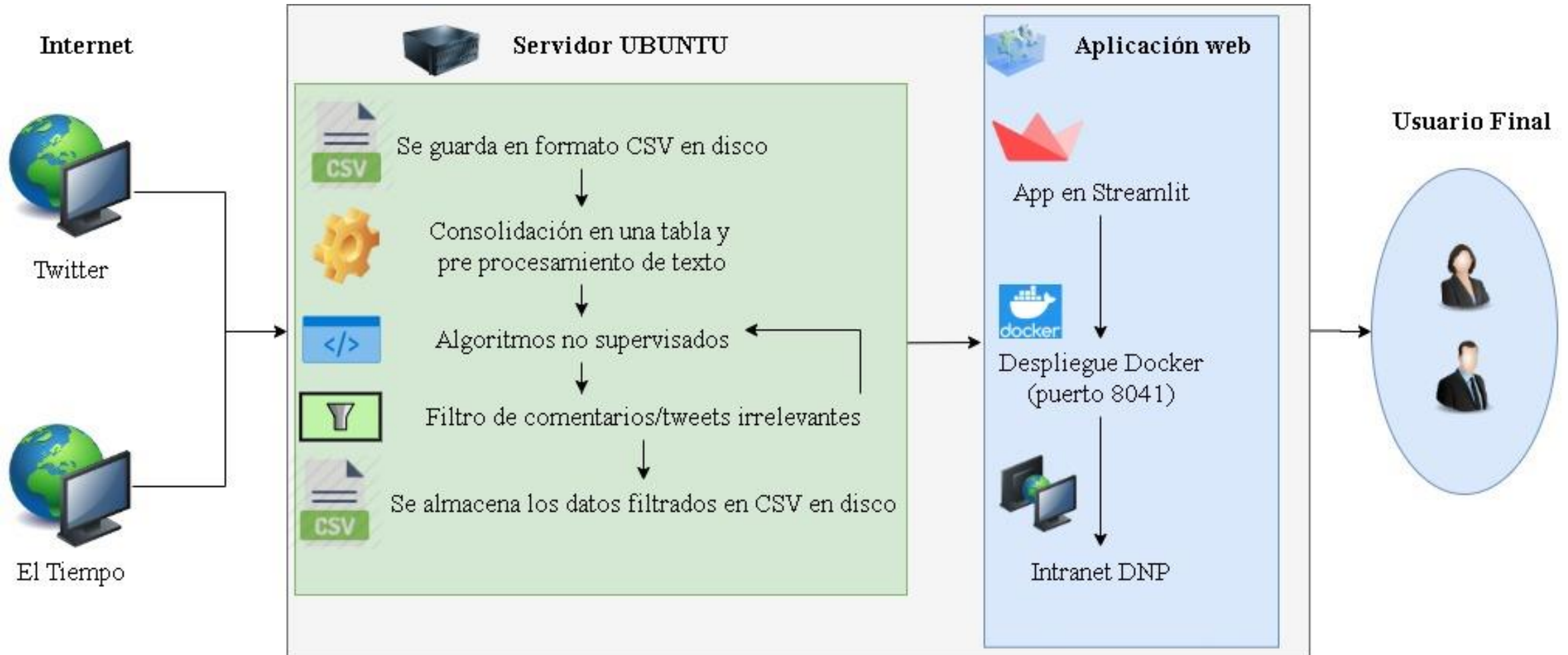
Identificación de sectores



TF-IDF



Aplicativo – Flujo de información



Conclusiones

1. Se desarrolló un aplicativo que permite entidades competentes saber de que hablan sobre temas relacionados sobre competencia desleal en la economía colombiana.
2. Se diseñó un flujo de información con base en modelos no supervisados que permiten realizar filtros de información con base en su temática
3. Es un caso práctico del uso de analítica en el sector público que sirve como insumo para la toma de decisiones de política pública en el país.



Gracias

Contacto:

crocha@dnpp.gov.co

ucd@dnpp.gov.co

www.linkedin.com/in/carlos-andrés-rocha-ruiz

