

# Herramienta para el análisis de difusión de noticias falsas en Twitter

Ignacio Laviña Faustmann

Doble Grado Ingeniería Informática y Administración y  
Dirección de Empresas (2019 – 2020)

Tutor: Antonio Berlanga de Jesús

Trabajo Fin de Grado



# Agenda

1

Objetivos y motivación

2

Definición del problema

3

La solución

4

Estado del arte

5

Contexto socio-económico

6

Contexto socio-económico

7

Marco regulador

8

Gestión del proyecto

9

Análisis del sistema

10

Diseño del sistema

11

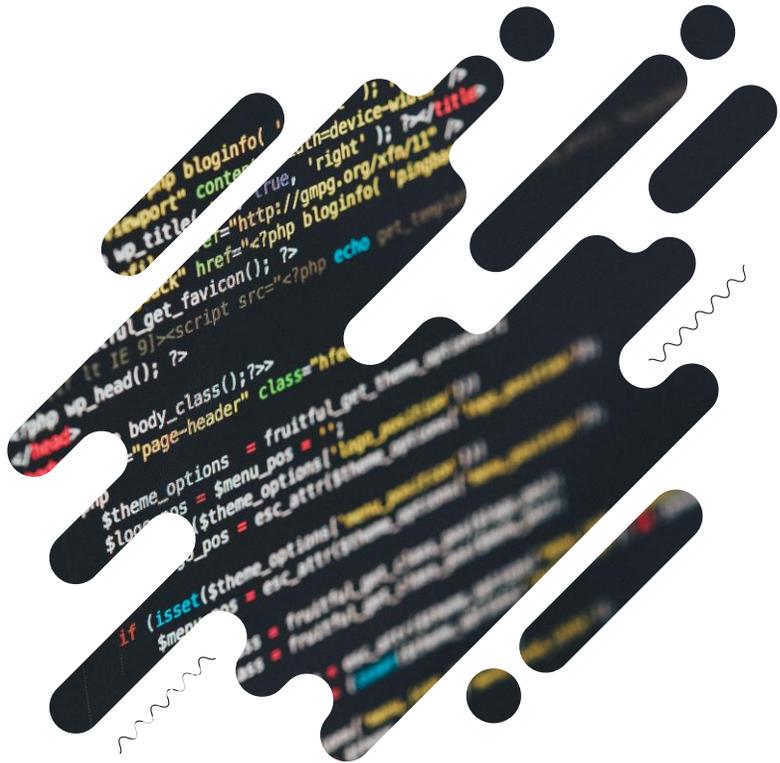
Resultados

12

Conclusiones



# Objetivos y motivación



## Crear una aplicación web

Crear una **aplicación web** real y funcional que sirva como **mínimo producto viable para validar** una idea.



## Generar un impacto positivo

Generar un **impacto positivo** en la Sociedad **ayudando** a aquellas personas y entidades que luchan contra las fake news.



## Demostrar los conocimientos

**Demostrar los conocimientos adquiridos** en diferentes asignaturas de la carrera, tanto en el aspecto teórico como en el aspecto práctico.



# El problema



## Noticias falsas

Se conoce como fake new o noticia falsa a aquellas noticias o elementos de información que no son veraces o que tienen tal grado de alteración que pueden considerarse como manipulaciones.

01

**+30% posibilidad de viralización [1]**

02

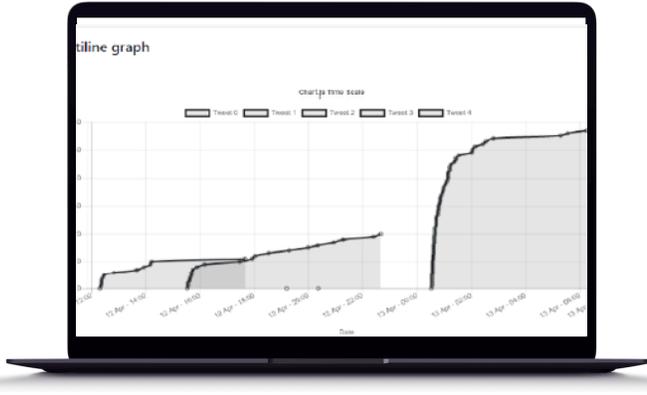
**Especial protagonismo de las redes sociales**

03

**Preocupación política y social**



# La solución



Una **aplicación web** que permita el **análisis de la difusión** de tweets y noticias en la red social **Twitter** y el seguimiento del alcance de las mismas de una forma sencilla o automatizada



Gestion de usuarios



Buscador de Tweets



Almacenamiento de Tweets



Seguimiento de cuentas relevantes



Análisis y visualización de datos



Agrupación de noticias similares



Contestación a través de la plataforma



Automatización



Probabilidad de cuentas bot



# Estado del arte

## Grandes tecnológicas

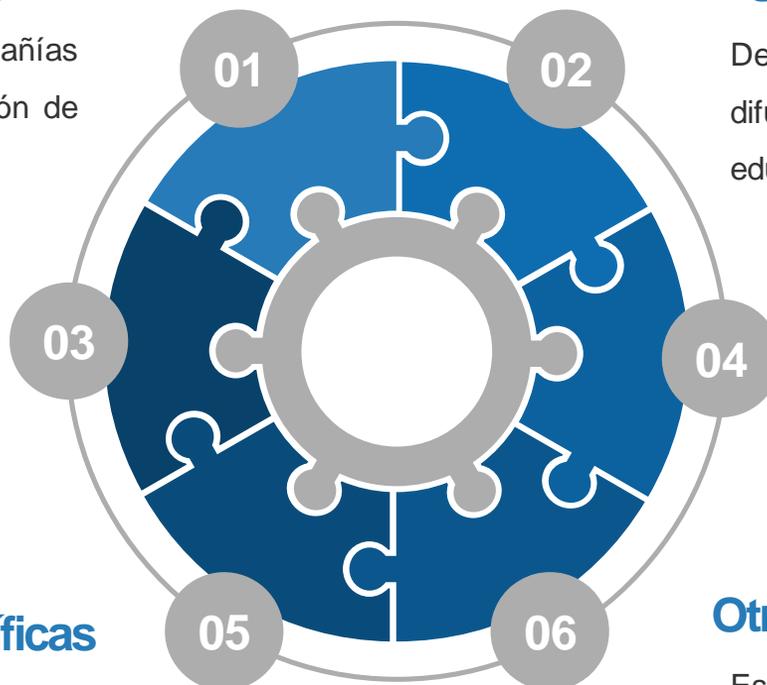
Utilización del algoritmos, adquisición de compañías estratégicas, uso de la comunidad de la red, creación de avisos y alertas y colaboraciones con agencias

## Herramientas de análisis

Visualización de datos y mejora de la capacidad de análisis

## Herramientas específicas

Plataformas Open Source dedicadas a analisis de cuentas falsas. Google fact check, Hoaxy



## Agencias de verificación

Dedicación específica a la verificación y el fact checking, difusión en medios sociales, labor de concienciación y educación. Ej: Newtral Malditobulo

## Detección de bots

Herramientas que usan Machine Learning y Procesamiento del Lenguaje para detectar bots

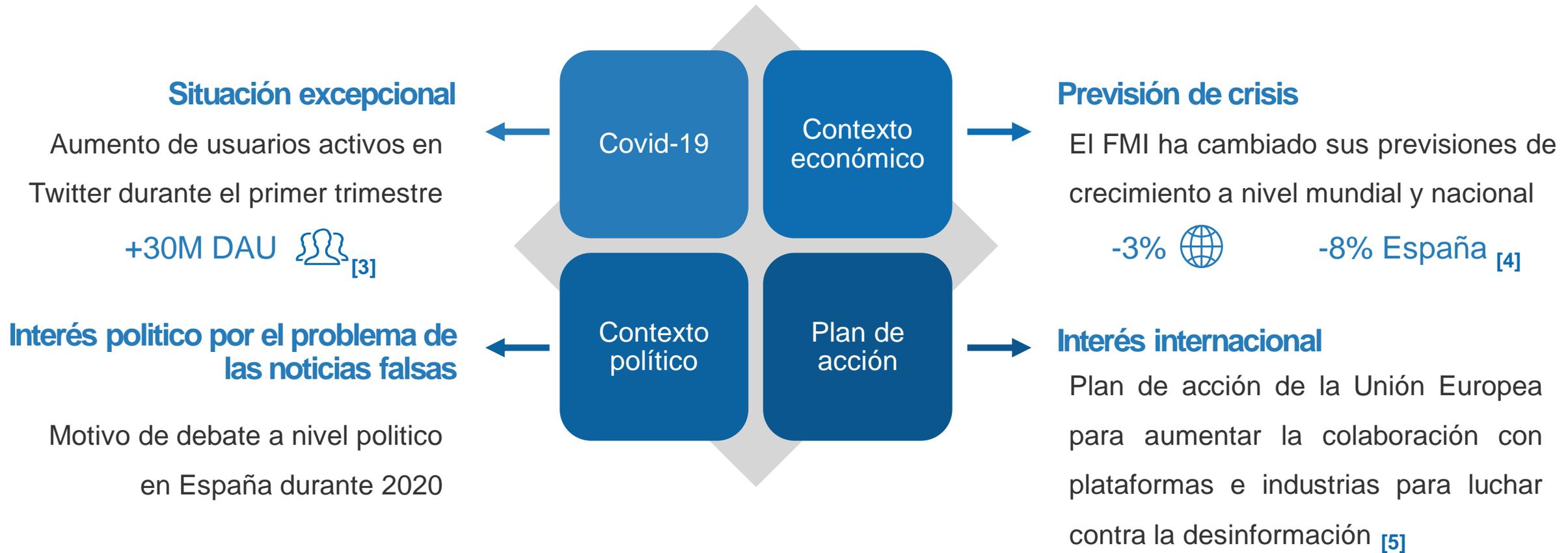
## Otras investigaciones y estudios

Estudios de la comunidad investigadora con diferentes enfoques relacionados con el Machine Learning, el Deep learning y el análisis de patrones.

Limitaciones con las fuentes de datos



# Contexto socio-económico



# GDPR

El Reglamento General de Protección de Datos (RGPD) consiste en el marco jurídico que regula la recogida, uso y tratamiento de los datos personales por parte de entidades y empresas con respecto a los ciudadanos europeos [5]

# Marco Regulator

01

Limitación del propósito de almacenamiento de los datos

02

Derecho al olvido y a la portabilidad de los datos

03

Consentimiento expreso para uso de datos con fines distintos a su uso legítimo

04

Control de riesgos y medidas de seguridad

# Marco Regulator

01

Los Tweets publicados son de dominio público

02

Está permitido la extracción de datos a través de la API

03

Limitaciones en el proceso de automatización de respuestas

04

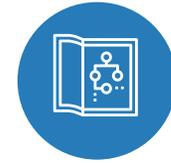
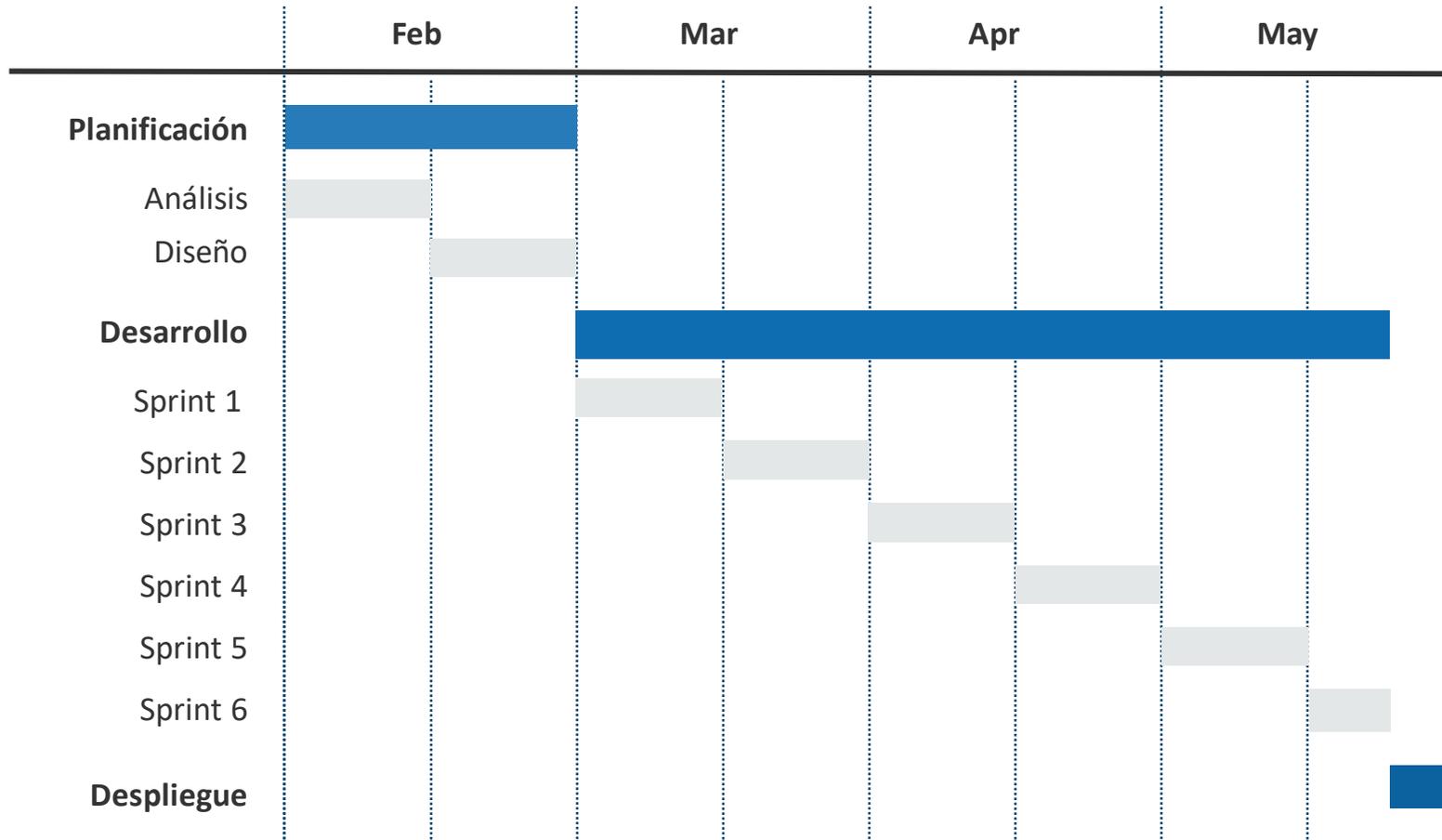
Actualización de status de un Tweet (eliminación o cambio de estado a privado)

## Twitter privacy policy

By publicly posting content when you Tweet, you are directing us to disclose that information as broadly as possible, including through our APIs, and directing those accessing the information through our APIs to do the same <sup>[6]</sup>

“

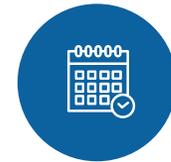
# Gestión del proyecto



Metodología ágil



Sprints de 15 días



Duración de 4 meses



# Gestión del proyecto

RECURSO	SALARIO BRUTO	% ASIGNACIÓN	HORAS DE PROYECTO	COSTES TOTALES
[Jefe de proyecto]	30.000€	30%	182	3.109€
[Arquitecto de software]	28.000€	100%	608	9.673€
[Desarrollador full stack]	28.000€	20%	122	1.935€
<b>[TOTAL]</b>				<b>14.716 €</b>

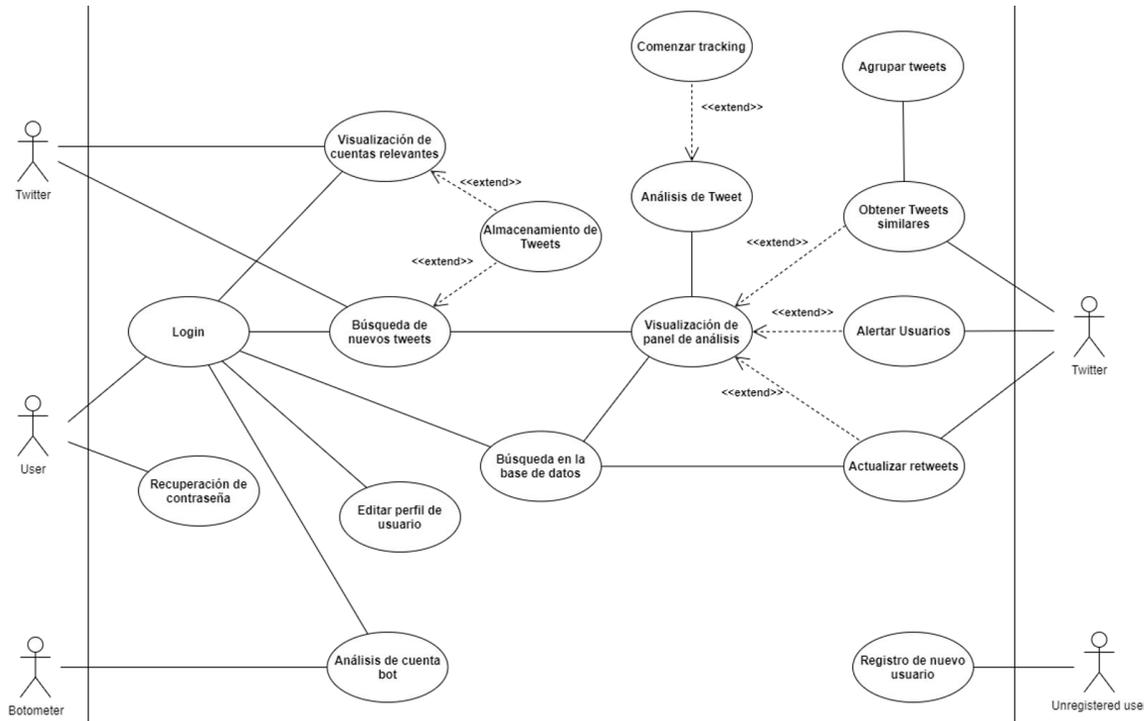
RECURSO	SALARIO BRUTO	AMORTIZACIÓN (MESES)	COSTES TOTALES
MacBook Pro	2.077 €	24	346 €
Dell Latitude	1.475 €	24	123 €
Dell Latitude	1.475 €	24	61 €
<b>[TOTAL]</b>			<b>531 €</b>

ROLE	RATE/HOUR
Sueldos y empleados	14.716 €
Equipos informáticos	531 €
Materiales fungibles	120 €
Software y licencias	200 €
<b>Subtotal</b>	<b>15.567 €</b>
Margen (10%)	1.156 €
<b>Total</b>	<b>17.124 €</b>

**TOTAL**  
**17.124 €**



# Casos de Uso



Identificación de los actores implicados en el sistema

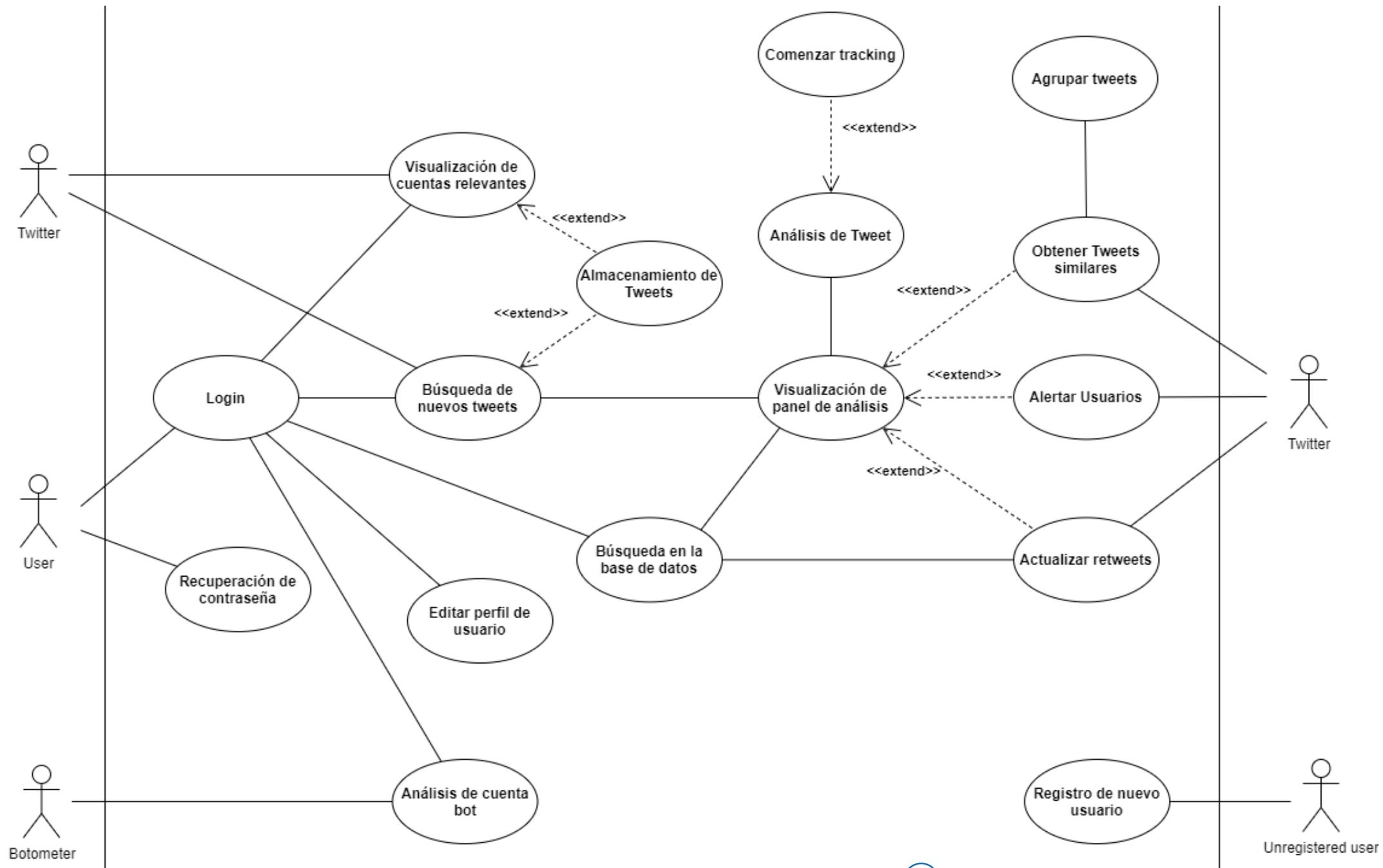


Elicitación de requisitos

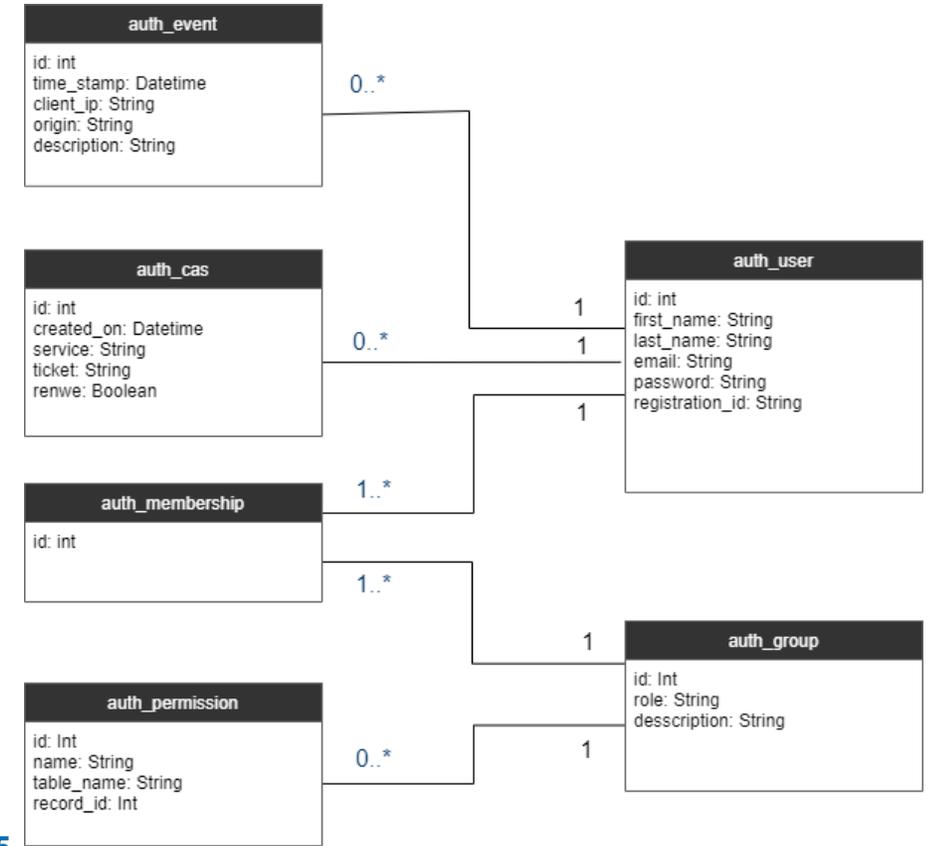
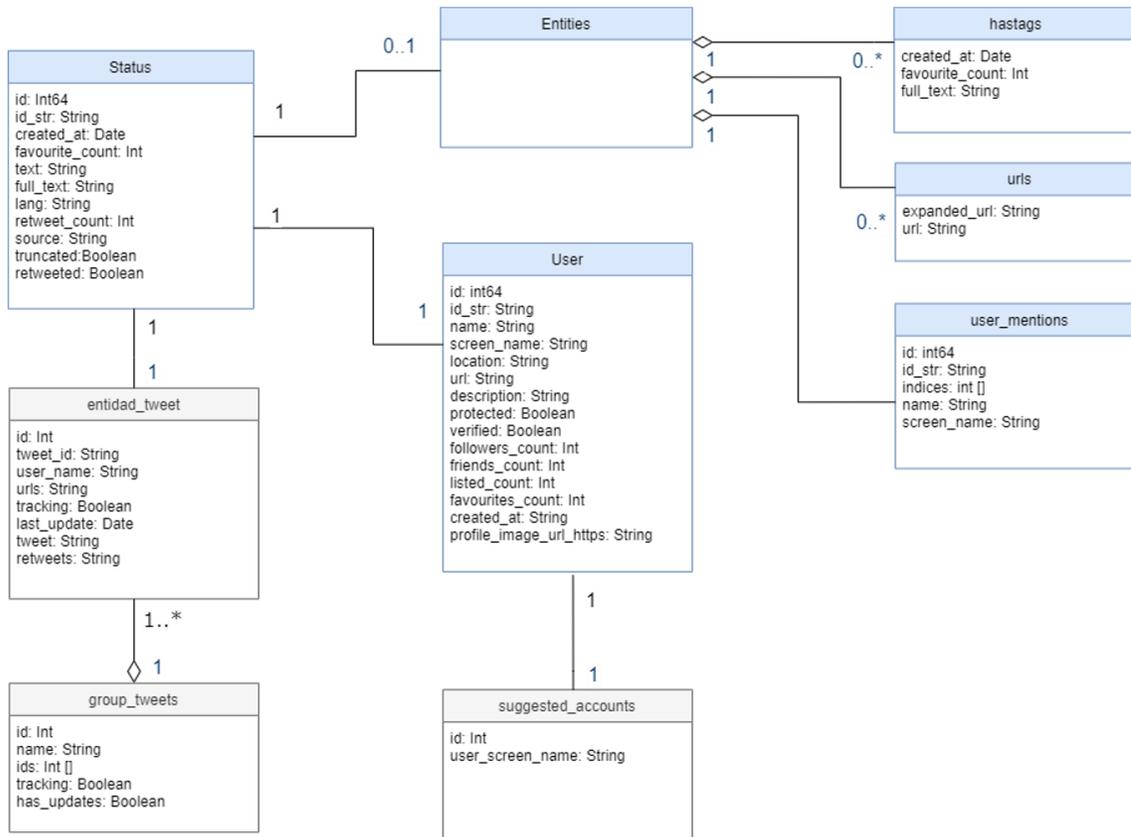


Facilita la planificación e implementación del proyecto

Caso de uso:	Nombre del caso de uso
<i>Identificador:</i>	Código de identificación alfanumérico secuencial: CU-XXX
<i>Actores</i>	Actores implicados en el caso de uso
<i>Descripción</i>	Descripción ampliada del caso de uso
<i>Prioridad</i>	Esencial / Deseable
<i>Referencias cruzadas</i>	Dependencias y relaciones con otros casos de uso



# Modelo de datos



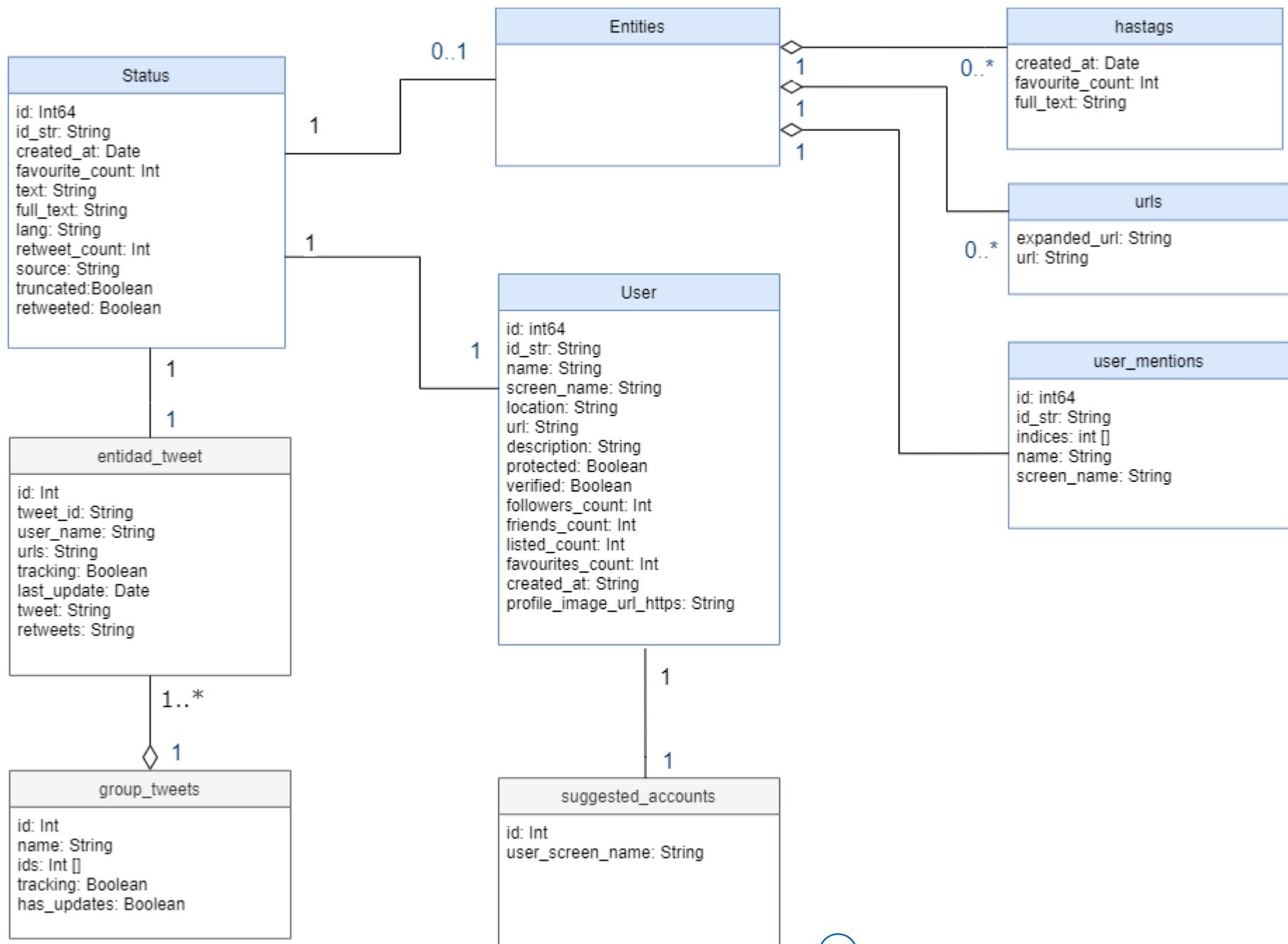
15

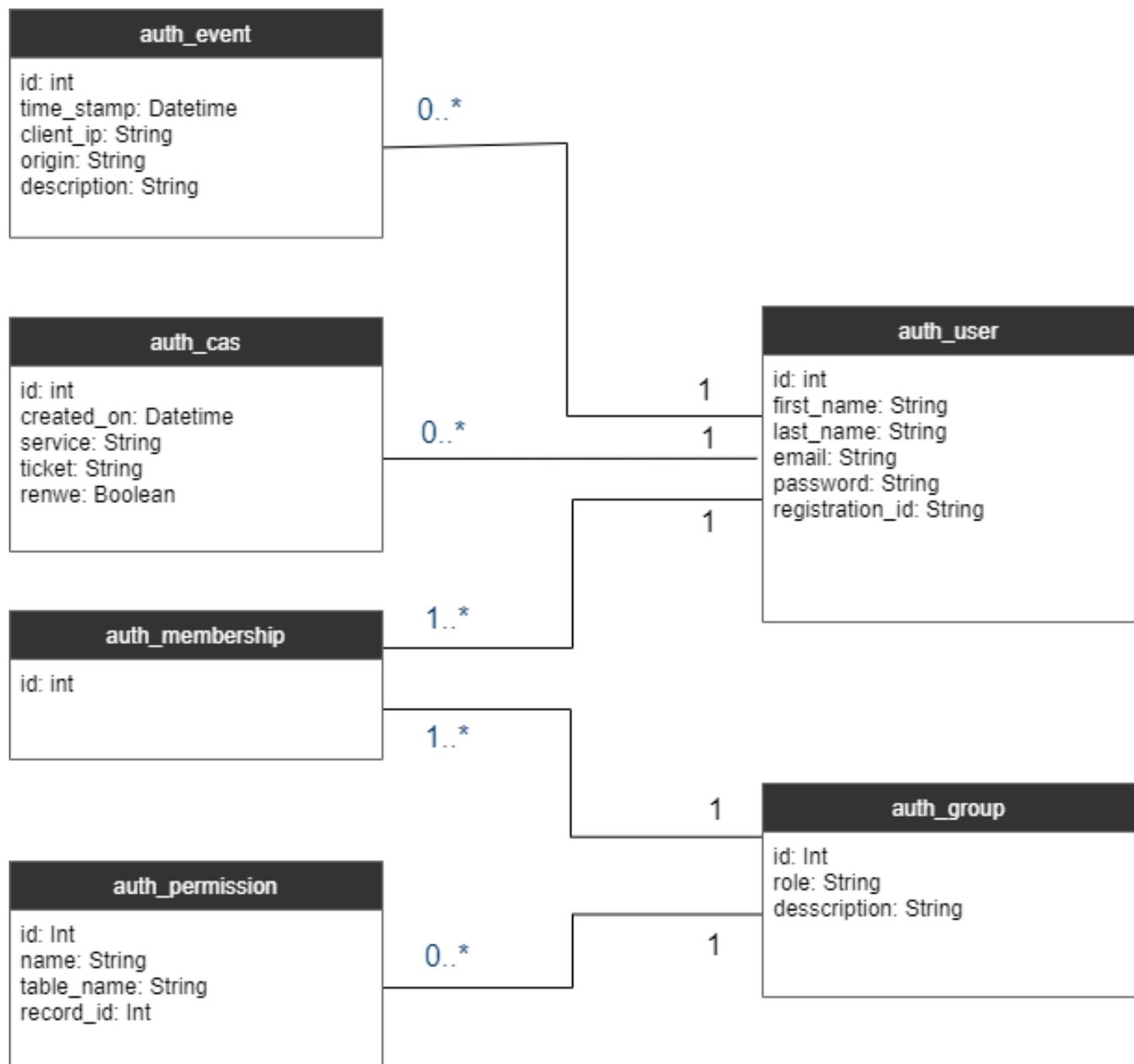
Modelo de datos Twitter

Modelo de datos de la aplicación

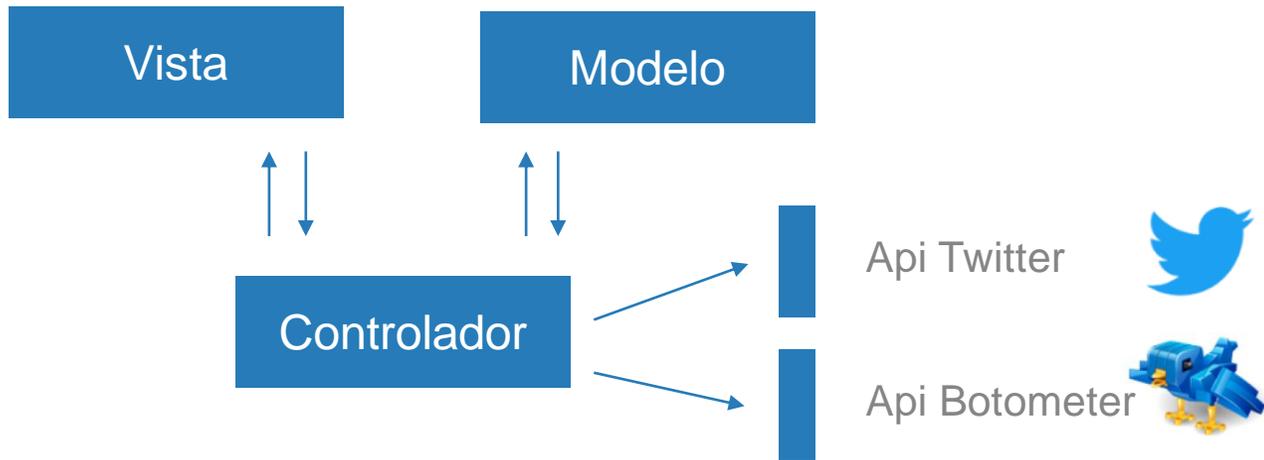
Modelo de datos Web2py







# Entorno operacional



## Web2py



Desarrollo express, con gestión de hilos y plantillas predefinidas



Gestión de usuarios



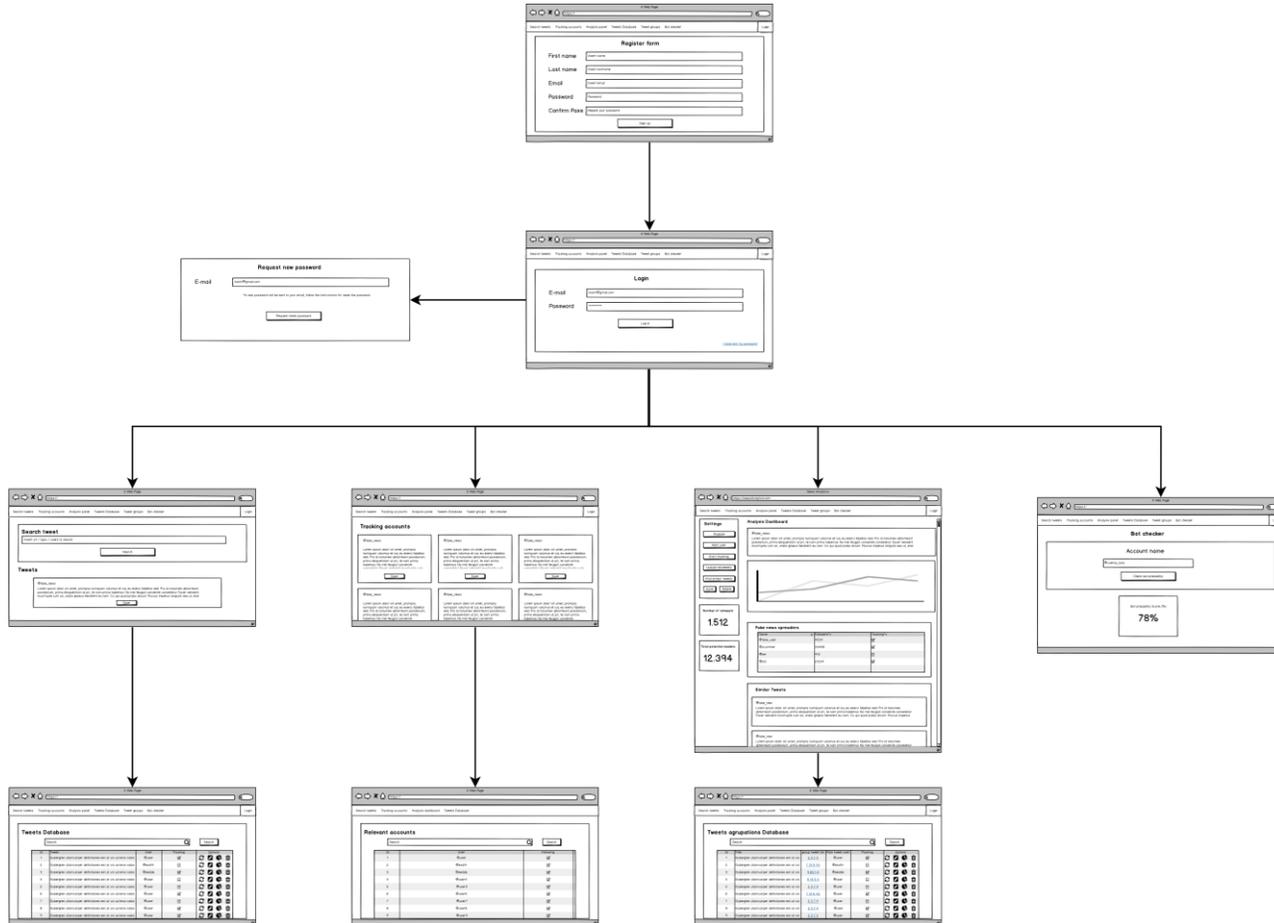
Seguridad testada frente a las principales vulnerabilidades



Base de datos integrada



# Diseño del sistema



News Analytics

https://newsAnalytics.com

Search tweets   Tracking accounts   Analysis panel   Tweets Database   Tweet groups   Bot checker   Login

### Settings

Analyze

Alert user

Start tracking

Update retweets

Find similar tweets

Save   Delete

---

Number of retweets

1.512

---

Total potential readers

12.394

### Analysis Dashboard

@fake\_news

Lorem ipsum dolor sit amet, prompta numquam volumus et ius, eu exerci fabellas sed. Pro id nonumes abhorreant posidonium, prima eloquentiam at pri, te nam prima habemus. No mei feugait convenire consetetur. Facer senserit incorrupte cum an, oratio graeco henderit eu nam. Cu qui quod probo dicunt. Mucius impetus singulis sea ut, erat

---

### Fake news spreaders

Name	Followers^v	Tracking^v
@fake_user	4034	<input checked="" type="checkbox"/>
@scammer	33458	<input checked="" type="checkbox"/>
@liar	413	<input type="checkbox"/>
@bot	21234	<input checked="" type="checkbox"/>

---

### Similar Tweets

@fake\_new

Lorem ipsum dolor sit amet, prompta numquam volumus et ius, eu exerci fabellas sed. Pro id nonumes abhorreant posidonium, prima eloquentiam at pri, te nam prima habemus. No mei feugait convenire consetetur. Facer senserit incorrupte cum an, oratio graeco henderit eu nam. Cu qui quod probo dicunt. Mucius impetus

---

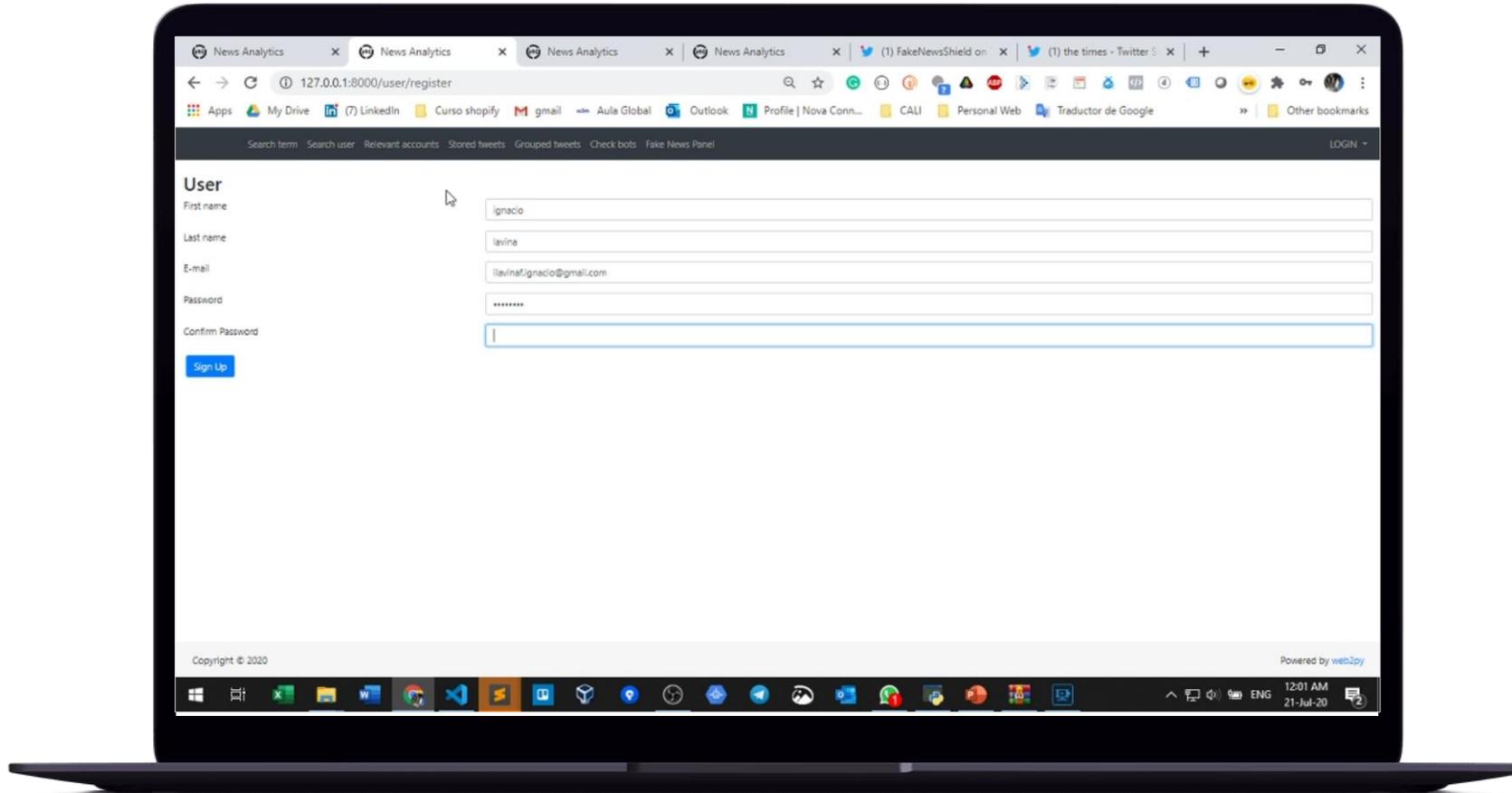
@fake\_new

Lorem ipsum dolor sit amet, prompta numquam volumus et ius, eu exerci fabellas sed. Pro id nonumes abhorreant posidonium, prima eloquentiam at pri, te nam prima habemus. No mei feugait convenire consetetur.

Herramienta para el análisis de diffusion de fake news en Twitter

18

# Resultados



## User

First name  
Last name  
E-mail  
Password  
Confirm Password

ignacio

lavina

ilavinaf.ignacio@gmail.com

\*\*\*\*\*

|

Sign Up

# Resultados

## Limitaciones

Versión gratuita de la API de Twitter

Dependencia tecnológica

Limitaciones legales

## Futuras vías de exploración

Uso de Inteligencia Artificial

Extrapolación a otras redes

Mejoras funcionales y visuales

## Explotación

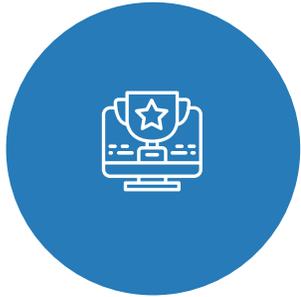
Software as a service. Modelo freemium

Servicio de información

Subvenciones y programas gubernamentales



# Conclusiones



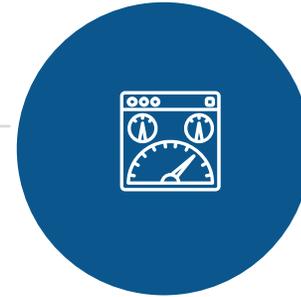
**Resultados**



**Conocimientos**



**Áreas de  
mejora**



**Impacto**

# Referencias

- [1] “Twitter privacy policy,” Twitter. [En línea]. Disponible en: <https://twitter.com/en/privacy>. [Último acceso: may 2020].
- [2] S. Aral, D. Roy, S. Vosoughi, “The spread of true and false news online,” *Science*, vol. 359, nº 6380, p. 7, 9 march 2018.
- [3] K. E. Matsa. E. Shearer, “News Use Across Social Media Platforms 2018,” *Pew Research Center*, 2018 [En línea].  
Disponible en: <https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/>.
- [4] A. Robinson, “El PIB de España caerá un 8% este año y el paro rozará el 21%, según el FMI,” *La Vanguardia*, 14 abril, 2020.
- [5] E. Schulze, “EU tells Facebook, Google and Twitter to take more action on fake news,” *CNBC*, 29 october, 2019.
- [6] European Union, “General data protection regulation, Regulation (EU) 2016/679,” 2016.
- [7] “Twitter privacy policy,” Twitter. [En línea]. Disponible en: <https://twitter.com/en/privacy>. [Último acceso: may 2020].



# Preguntas?



## User

First name  
Last name  
E-mail  
Password  
Confirm Password

ignacio

lavina

ilavinaf.ignacio@gmail.com

\*\*\*\*\*

|

Sign Up