SentiUMU+: *Análisis Avanzado del Sentimiento Digital en Redes Sociales*

Sistema avanzado de monitorización y análisis de la percepción universitaria en redes sociales

---

Javier García Ros <*jgarcia@um.es*>, *Universidad de Murcia* Antonio Arques Acosta <*antonio.arques@ticarum.es*>, *Ticarum*

21 May 2025

UNIVERSIDAD DE MURCIA

https://www.um.es

Agenda

1. **Antecedentes**
2. **Objetivo**
3. **Conceptos**
4. **Fuentes de Datos**
5. **Uso de IAs generativas para análisis**
6. **Medodología**
7. **Paneles**
8. **Demo**
9. **Conclusiones**

Figura 1: Sentiumu+

Análisis de Sentimiento en RRSS

El análisis de sentimiento es una técnica de **procesamiento de lenguaje natural (NLP)** que permite identificar y extraer opiniones y emociones de un texto.

Ejemplo Tweet real

> *"Eso de que en la UMU a partir de las 14h es como si la vida se parase y nadie te coja el teléfono me tiene sobrecogida. Pero, sin duda, lo que más me preocupa es que antes de esa hora, TAMPOCO TE RESUELVEN NADA (si tienes la suerte de que te atiendan, claro)"* Irene de las flores, ***Oct 22, 2024***

Análisis

- **Sentimiento**: *Muy Negativo*

Importancia del Análisis de Sentimiento

Comprensión del Alumno/Usuarios:

- Ayuda a *entender* las opiniones de nuestros alumnos/usuarios.

Monitoreo de imagen:

- Nos permite *rastrear* la percepción pública.
- Representarla gráficamente
- Ver la evolución en el tiempo
- Cuantificar valores

Detección de crisis:

- Nos permite *crear alertas automáticas* y responder en menos tiempo ante situaciones de crisis.

Toma de Decisiones:

- Facilita decisiones informadas basadas en datos.

Antecedentes Análisis de Sentimiento

Año 2019 Prueba concepto

- Extracción de tweets (twitter) mediante **API**
  - mucho más sencillo antes que ahora
- análisis de sentimiento técnicas:
  - *machine learning*
  - Usando **R** y **Tidytext** de *Julia Silge*
  - No había buen soporte en **Español**
  - Traducción herramientas **en linea** (apis)
- creación informe estático

Rediris 2019

- Charla relámpago: **Análisis de sentimiento basado en tweets en la Universidad de Murcia**

Concepto LLM

Irrupción de los LLM Large Language Model

Los **LLM** o *Modelos de Lenguaje de Gran Tamaño* son:

- modelos entrenados con conjunto de datos masivos
- reconocen, traducen y generan texto
- imitan el lenguaje natural
- gran salto de los modelos tipo **ChatGPT**

Ejemplos

- Cerrados como:
  - *Chapgtp*
  - *Gemini* ...
- Abiertos como:
  - *Llama*
  - *Gemma*
  - *Deepseek* ...

Objetivo (I)

Alcance Inicial Proyecto 2024

- Análisis y seguimiento del **sentimiento** en **Twitter** a lo largo del tiempo
  - Extracción de tweets
  - Mostrar resultados paneles de **Lagar**
    - Infraestructura nuestra con **ELK**
  - Tiempo real

Objetivo (II)

Multi-idioma (no sólo en español)

- Mayoritarios con muy buen soporte
- **Chatgpt 4.0** (unos 100 idiomas)
- **Gemma 3.0** (140 idiomas)

Arquitectura Multi Red Social

- **Bluesky**
- Posibles incluir más: *Threads*, *Mastodom*, ...

Arquitectura Multimedia

- ¿Podemos analizar algo más que texto? *Imágen, audio, video...*
- ¿**Instagram**?,...

Objetivo (III)

Ir ganando experiencia con los LLM´s

- **local** (no tenemos disponibles *GPU´*s):
  - Cada vez modelos más interesantes
  - Con pocos requisitos
- **Alternativas cloud**:
  - Azure, ...

Acceso a las publicaciones

- ~~**Tiempo real**~~
- Lotes

Objetivo (IV)

Algo más que sentimiento

- Análisis **multinivel**:
  - **sentimiento**
  - identificar entidades descritas
  - mejor análisis de la relevancia
  - análisis tono
  - idioma
  - recomendaciones automáticas
  - cuantificar valores
  - etc.

Integración herramientas visualización/monitorización

- Visualización gráfica mediante paneles (**Kibana**)
- Posibilidad de definir **KPIs** de seguimiento
- Generación automática de **alertas**
- ...

Ejemplo Tweet real con LLM

Campo	Valor
Nombre	*Jonathan Lopez-Vera*
Red Social	*BlueSky*
Idioma	*Catalán*
Contentido	*Uns segons de la meva xerrada ahir a la Universidad de Murcia, on m'han tractat fantàsticament bé. Moltes gràcies per convidar-me a venir!*
Categoria	**Positivo**
Descripción	El texto refleja un sentimiento muy positivo, el autor expresa gratitud y satisfacción por el trato recibido en la Universidad de Murcia.
Relevancia	**Muy Relacionado**
Confianza	0.88
Recomendaciones	Continuar brindando un trato excepcional a visitantes y ponentes. Promover y destacar las experiencias positivas en la universidad en redes sociales. Agradecer públicamente a los ponentes por su participación.

Metodología

Metodología

1.) Extracción de los tweets diarios: ¡sin usar API!

2.) Preprocesamiento: Limpieza y normalización

3.) Análisis mediante técnicas de IA (multi-técnica):

- Sentimiento
- Entidades
- Temas
- ...

4.) Inyección de logs enriquecidos en ELK/LAGAR

5.) Paneles de visualización ELK/LAGAR

6.) Posibilidad de conectar con Alertas (pendiente)

Metodología

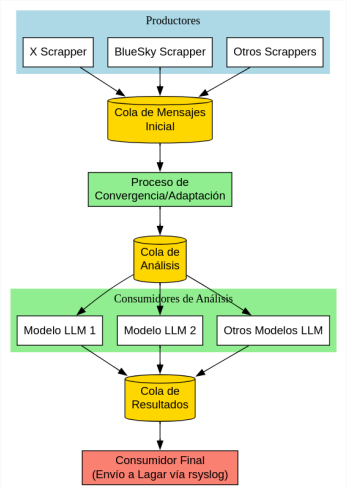


Figura 2: Arquitectura

Arquitectura basada en Colas

Multiples colas **Kafka** con diferentes grupos consumidores

- **Ventajas**:
  - Asíncrono
  - Escalable
  - Versátil: Facilidad de inclusión de nuevas fuentes, funciones..
  - Tratamientos de datos paralelos / Multitarea
  - Robustez

Metodología (I)

Extracción de Posts

Generalmente dos tipos de fuentes de datos:

Uso de APIs / Métodos tradicionales

- Documentación extensa
- Simplicidad de Uso
- Con el auge de modelos voraces de datos:
  - Limitaciones en uso por posts
  - Barreras de Pago

Técnicas de Scrapping

- Generalmente *Best Effort*
- Basadas en emular navegadores web (comportamiento *"humano"*)
- Complejidad de Implementación
- Limitaciones de Uso (IP Ban)

Metodología (II)

Análisis de Sentimientos

- Modelos de IA generativa pesados
- Modelos en Nube Vs Modelos en Local
- Pieza clave **Robustez**
  - Es imprescindible asegurar el comportamiento y formato
  - Diversas técnicas:

Técnicas:

Prompting

Refuerzo con Ejemplos

Modelos IA evaluadores de formato

Otras Técnicas
Transfer learning
Fine Tuning

Metodología (III)

Almacenamiento de Datos

- Envió de resultados a **ELK** mediante **RSyslog**.
- Almacenamiento en Índices en **ElasticSearch**
- Representación de Métricas en Cuadros de Mando (**Kibana**).

Visualizaciones

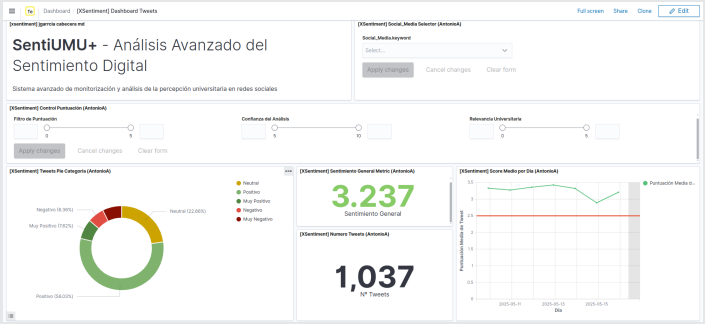


Figura 3: Selección de Visualizaciones

Metodología (IV). Mejoras

Análisis Multimedia

- Análisis de formatos de **audio** y **video**
  - Uso de modelos de transcripción y análisis de video
  - Mayor coste computacional y/o económico
- El diseño basado en colas de mensajes permite una incorporación simple: nuevos productores de datos.

Evaluación del Proceso

- Nuevo consumidor: IA generativa con la tarea de evaluar el análisis realizado
  - Es más simple evaluar una tarea que crearla
  - Feedback sobre el funcionamiento

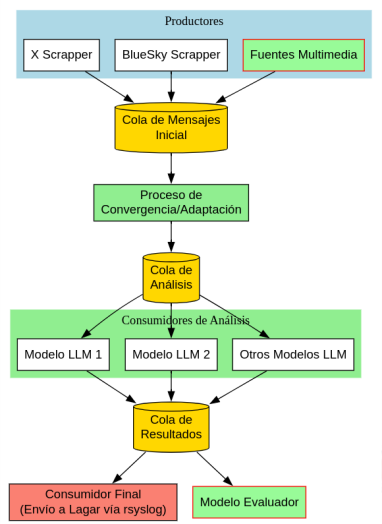


Figura 4: Mejoras

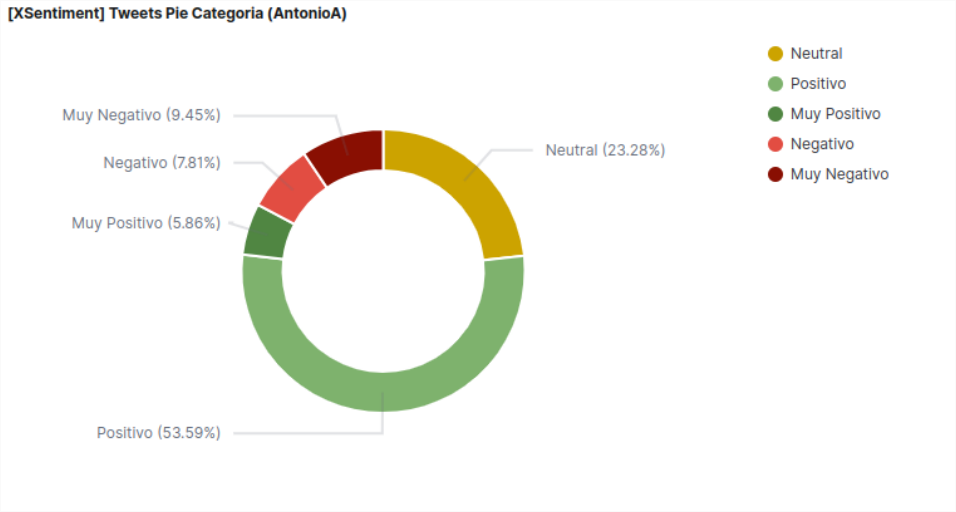
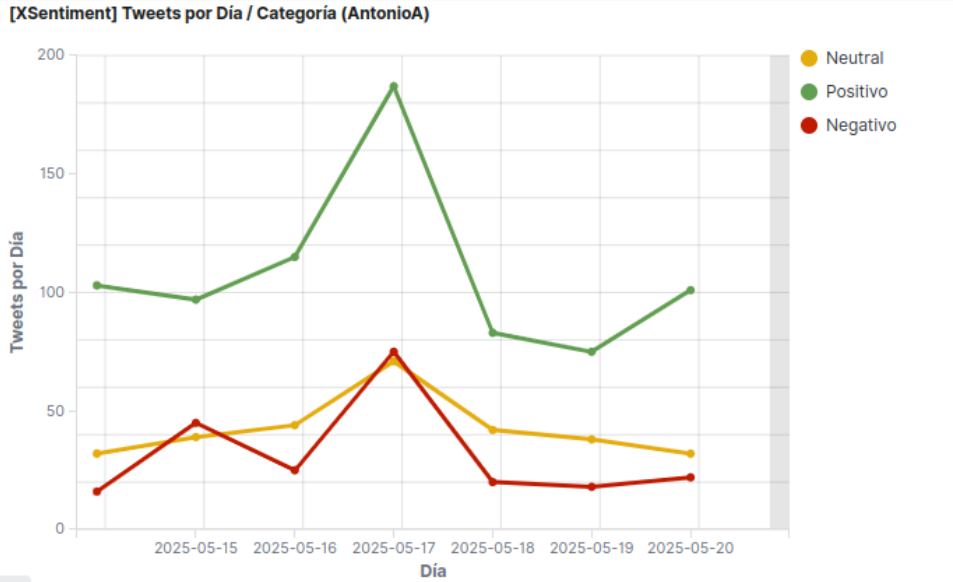


Figura 5: Selección de Visualizaciones

Dashboards (II)



Dashboards (II)

Tweet	Usuario	Recomendación	Temática	Fecha	Puntuación Sentimiento
📌 Próxima cita en Meta Librería de Madrid Os esperamos el lunes 13 de enero para la presentación del MANUAL DE ARGUMENTACIÓN. Con la participación del autor @jagascon.bsky.social (Universidad de Murcia) y Hubert Marraud (Universidad Autónoma de Madrid)	plazayvaldes.bsky.social	Considerar la grabación o transmisión en vivo del evento para llegar a una audiencia más amplia	Autores universitarios	Jan 2, 2025 @ 21:14:41.347	3.8
Del 8 al 10 de julio se celebró en el Museo Arqueológico Municipal de Lorca una nueva edición del curso de Arqueología Medieval de Sefarad organizado por la Universidad de Murcia y el ayuntamiento de Lorca. El curso, d... https://s.mtrbio.com/bouzgrfruz	blancosobrenegro.bsky.social	Ampliar la difusión de estas actividades para involucrar a más personas interesadas en la arqueología medieval.	Colaboración institucional	Jan 7, 2025 @ 21:57:21.405	3.8
🔍 INVESTIGACIÓN	La Universidad de Murcia aprueba nuevas políticas de RRHH en investigación alineadas con directrices europeas 🇪🇺✨ ➕🔗 https://t.um.es/otm-r	um.es	Comunicar de manera efectiva los beneficios y detalles de estas nuevas políticas a la comunidad universitaria.	Internacionalización	Jan 8, 2025 @ 10:02:54.679	3.7
PREMIO 🏆	El artista Telmo Meana ha ganado el primer premio del XXIII Concurso de Fotografía de la Universidad de Murcia por su obra Autorretrato como cadáver exquisito 📸🏆 ➕🔗 https://t.um.es/Telmo-Meana	um.es	Destacar más eventos culturales realizados por la universidad para fortalecer su conexión con el arte local.	Actividades culturales universitarias	Jan 11, 2025 @ 11:01:41.362	3.8
El cambio climático "aumentará la frecuencia de las danas", según un estudio de la Universidad de Murcia www.elespanol.com/enclave-ods/...	troposfera.bsky.social	Impulsar la colaboración interuniversitaria en investigaciones relacionadas con el medio ambiente.	Investigación sobre cambio climático	Jan 15, 2025 @ 20:10:17.547	3.5
el #proyectoumu es mi proyecto más ambicioso hasta la fecha: -dark academia -rivales académicas -universidad de Murcia en 1897 -estudio de la poesía clásica -se pueden hacer conjuros con poemas 👀	neonatasaurus.bsky.social	Explorar la posibilidad de realizar eventos o talleres sobre 'dark academia' y su relación con la poesía.	Cultura universitaria	Jan 21, 2025 @ 18:40:19.936	3.8
Hoy, taller de traducción de cómic en la Universidad de Murcia. 💭💥	scheherezade.bsky.social	Difundir ampliamente eventos como este taller para aumentar la participación estudiantil.	Actividades extracurriculares	Jan 22, 2025 @ 12:15:04.538	4.2

Figura 7: Selección de Visualizaciones

Conclusiones

Gran potencia y versatilidad de modelo basado en colas

- **Kafka** nos ha permitido por alguna de sus especiales características implementar muchas mejoras de forma sencilla.
- Modelo muy recomendable para otras aplicaciones.

Facilidad de ampliación de Fuentes y tipo de Datos

- Relativa facilidad para incluir otras redes sociales como **Mastodom** o **Threads** e incluso en un o más basados en video como **Instagram** enriquecerá el análisis y ofrecerá una visión más completa.

Modelos Pequeños / Modelos Pesados

- Los modelos abiertos pequeños (**qwen2 0.5b**, **llama3 8b**, ...) ofrecen mejores tiempos de computo y menor coste, pero pierden robustez de formato y semántica. Aunque este tipo de modelos es cada vez mas competente (**Deepseek-r1 1.5b**, **Microsoft Phi 4.0 14b**,).

Conclusiones (ii)

Importancia del Análisis de Sentimiento

- Es esencial para entender las opiniones de los usuarios y tomar decisiones informadas.

Capacidad de cuantificar y crear métricas

- De una manera más fácilmente reproducible.
- También posibilidad de uso interno en otros entornos:
  - como evaluación de Jira no sólo como los 5 iconos.
  - preguntas abiertas encuestas.
  - ...

Conclusiones (iii)

Protección frente a las IAs

- El acceso a los datos se está complicando como una misma defensa a la extracción masiva/abuso

Conveniencias de incluir LLM más pequeños/económicos

- Evaluar es menos costoso que crear.
- Puedes usar LLM´s más sencillos para evaluar algunos resultados (muestreo)
- Por ejemplo: no todos los LLMs generan buenos datos estructurados.

Engrasar el uso de LLM para tareas mayor impacto en areas de negocio

- Mesa de trabajo para probar distintas tecnologías e ideas

¡¡MUCHAS GRACIAS!!

*"La sociedad necesita todo tipo de habilidades que **no son sólo cognitivas**, son **emocionales**, son **afectivas**. No podemos montar la sociedad sobre datos"* ***Alvin Toffler***

¿Preguntas?