Análisis de Desinformación en el Sector Bancario

Un enfoque basado en métricas temporales y análisis de narrativas

Octubre 2024

Metodología

- **Datos de entrada:**
  - 200K tweets
  - 866 clústeres
  - Período: Octubre 2024
- **Sistema de scoring:**

$$S_c = \tau_c \cdot \rho_c \cdot \mu_c$$

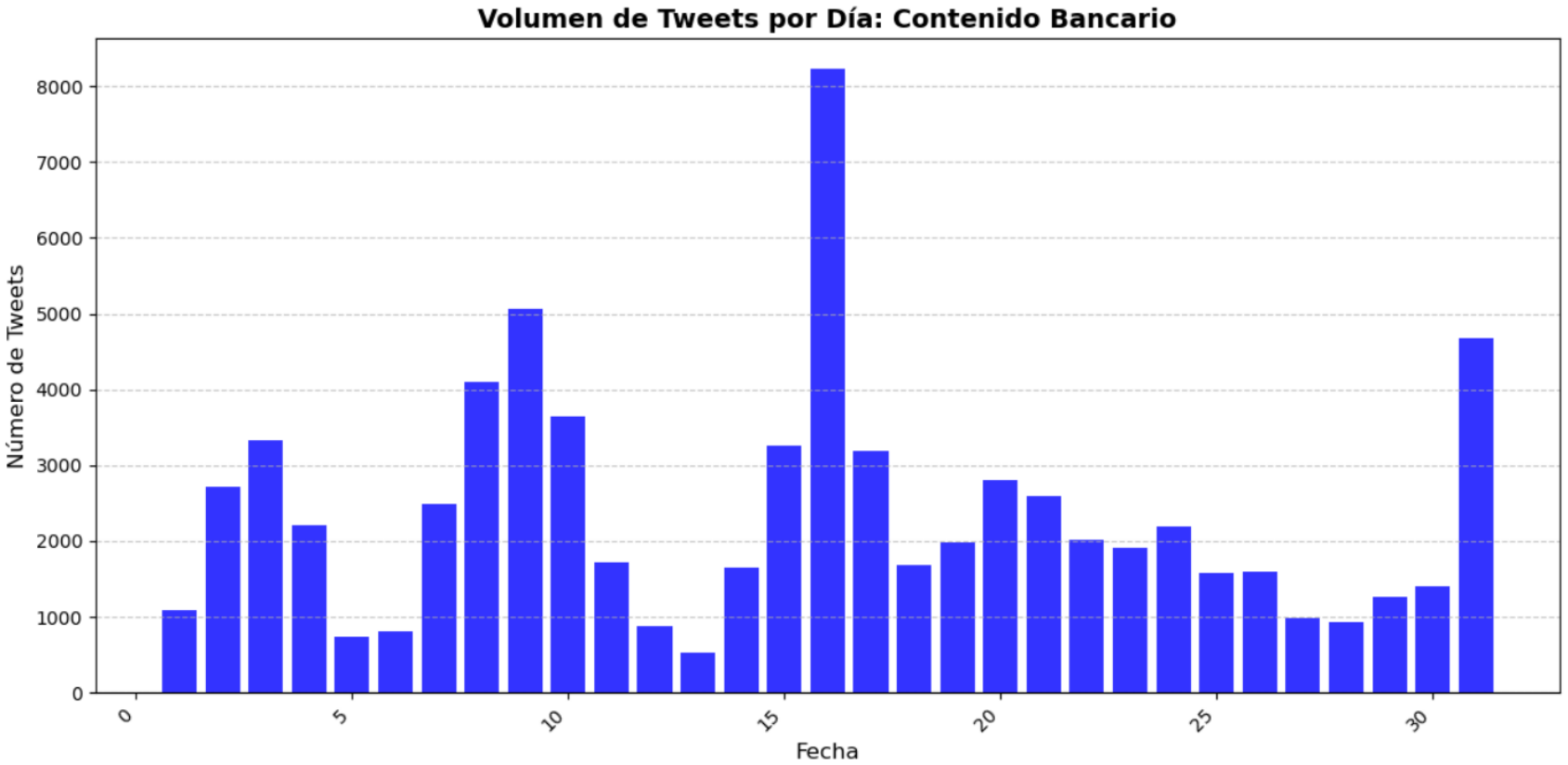
donde:

- τ_c: ratio de toxicidad
- ρ_c: ratio temporal
- μ_c: ratio de usuarios únicos

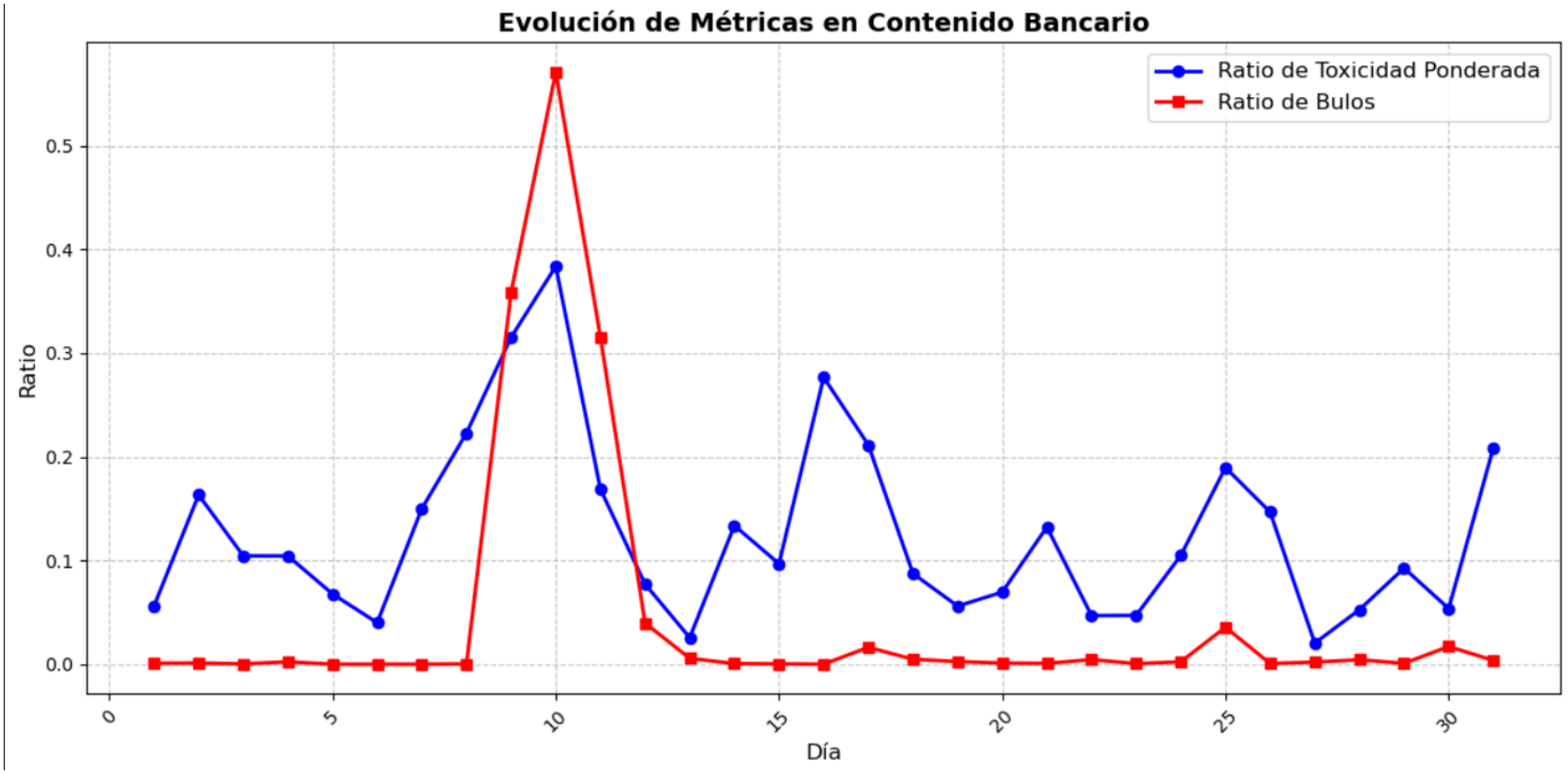
Métricas Fundamentales

- **Ratio de toxicidad ponderada (S_c)**
  - Medida compuesta [0,1]
  - Pondera toxicidad, alcance y dispersión
- **Patrones de propagación**
  - Ratio retweets/originales
  - Distribución temporal
  - Concentración de usuarios
- **Detección de bulos**
  - Verificación fact-checkers
  - Correlación con picos de S_c

Volmetría tuits bancarios



Toxicidad ponderada y Fact-checking



- Pico máximo: 10 octubre
- Correlación con máximo de S_c
- Períodos de baja detección requieren análisis manual

Franjas Temporales Críticas

- **2-4 octubre:** Inicio de narrativas bancarias
- **8-10 octubre:** Máximo histórico en S_c y detección de bulos
- **15-17 octubre:** Pico en volumen de contenido bancario
- **24-26 octubre:** Alto S_c con volumen moderado
- **31 octubre:** Repunte final

Hallazgos Principales

Alta probabilidad de bulo:

- Declaraciones Lagarde
- Narrativa xenófoba bancaria

Contenido manipulado:

- Financiación armamento
- Presiones políticas

Criterios de Identificación

- **Señales de desinformación:**
  - Alta ratio retweet/original ($> 1000 : 1$)
  - Distribución temporal artificial
  - Concentración de usuarios similares
  - Contenido emotivo/polarizante
- **Señales de veracidad:**
  - Fuentes verificables
  - Distribución temporal natural
  - Diversidad de usuarios
  - Base en hechos comprobables

Conclusiones

- La detección temprana de bulos es crucial (caso Lagarde)
- Los patrones temporales son indicadores clave
- La métrica S_c es efectiva para identificar contenido sospechoso
- La combinación de métricas cuantitativas y análisis narrativo mejora la detección

Recomendación: Implementar sistema de alerta temprana basado en S_c y patrones de propagación

Sistema de Alerta Temprana: Forecasting Multivariante

Arquitectura Propuesta

- **LSTM Bidireccional**
  - Captura dependencias temporales largas
  - Analiza secuencias en ambas direcciones
- **Transformer con atención temporal**
  - Identifica correlaciones entre métricas
  - Pondera importancia de eventos pasados

Variables de entrada

- Series temporales de S_c y componentes
- Patrones de propagación (retweets/hora)
- Métricas de usuarios (concentración/dispersión)
- Indicadores de toxicidad granular τ'_c

Sistema de Predicción y Alertas

Predicciones del Modelo

$$\hat{S}_c(t+\Delta) = f(\{S_c(t-k)\}_{k=0}^{n}, \{\tau'_c(t-k)\}_{k=0}^{n}, \{\text{features}(t-k)\}_{k=0}^{n})$$

donde $\Delta \in \{24h, 48h\}$ y n es la ventana temporal

Métricas de Alerta

- $P(\text{campaña}|S_c, \tau'_c)$
- $A(t) = \|\hat{S}_c(t) - \mathbb{E}[S_c]\|$
- $\text{Umbral}_t = \mu_t + k\sigma_t$

Sistema de Alarmas

- Crítica: $P > 0{,}8$
- Alta: $P \in [0{,}6, 0{,}8]$
- Media: $P \in [0{,}4, 0{,}6]$
- Baja: $P < 0{,}4$

Sistema de Predicción y Alertas

Beneficios

- Detección temprana de campañas coordinadas
- Umbrales adaptativos según contexto temporal
- Cuantificación de incertidumbre en predicciones
- Priorización automática de investigación

Margen de mejora: Sistema de Scoring Extendido

- **Versión actual:** Multiplicativa

$$S_c = \tau_c \cdot \rho_c \cdot \mu_c$$

- **Versión extendida:** Combinación lineal convexa

$$S'_c = \alpha\tau_c + \beta\rho_c + \gamma\mu_c$$

donde $\alpha + \beta + \gamma = 1$

Ventajas de S'_c

- Ajuste flexible (para cliente) de pesos según contexto
- Mayor granularidad en la detección
- Permite priorizar factores específicos
- Calibración basada en retroalimentación

Margen de mejora: Sistema de Scoring Extendido

Ejemplo

$S'_c = 0{,}5\tau_c + 0{,}3\mu_c + 0{,}2\rho_c$

- Prioriza toxicidad (50 %)
- Enfatiza dispersión de usuarios (30 %)
- Menor peso al factor temporal (20 %)

Granularidad en la Toxicidad

$$\tau'_c = \sum_{i \in T} \omega_i s_i + \sum_{j \in O} \lambda_j o_j$$

Componentes ofensivos (T):

- Sarcasmo
- Amenaza
- Insulto
- Agresión

Objetivos (O):

- Keywords bancarias
- Entidades específicas
- Productos financieros
- Servicios bancarios

Ejemplo de distribución de pesos

- $\lambda_{\text{banco}} = 0{,}4$ (keywords bancarias)
- $\omega_{\text{sarcasmo}} = 0{,}3$ (componente más correlacionado)
- $\omega_{\text{amenaza}} = 0{,}15$
- $\omega_{\text{insulto}} = 0{,}15$

Granularidad en la Toxicidad

Beneficios

- Mayor precisión en detección
- Identificación de patrones específicos
- Adaptabilidad al contexto bancario
- Índice continuo $\tau'_c \in [0, 1]$