Tarea 3: Detección de Bots con BERT y NLP

Curso: Fundamentos de Inteligencia Artificial

Duración estimada: 2 sesiones (3 horas)

---

1. Objetivo general

Desarrollar una aplicación práctica en **Streamlit** que permita detectar posibles cuentas automatizadas (bots) en una red social de libre elección, utilizando técnicas de **procesamiento de lenguaje natural (NLP)** con **BERT** y un modelo de detección de anomalías basado en comportamiento y contenido textual.

2. Competencias específicas

- Aplicar modelos de lenguaje preentrenados (BERT) para obtener representaciones semánticas de texto.
- Integrar variables de comportamiento, léxicas y semánticas en un pipeline de análisis.
- Desarrollar una aplicación web interactiva para visualizar resultados y justificar hallazgos.
- Reflexionar sobre el uso ético de modelos de detección automática en redes sociales.

3. Descripción del taller

Cada estudiante o grupo seleccionará una red social de interés (por ejemplo, Twitter/X, Reddit, YouTube, Instagram o Mastodon) y recopilará un conjunto de publicaciones públicas respetando los Términos de Servicio y la privacidad de los usuarios. El conjunto de datos deberá incluir las columnas:

Columna	**Tipo**	**Descripción**
user_id	texto	Identificador del usuario (anonimizado)
text	texto	Publicación del usuario
timestamp	fecha	Fecha/hora del post
likes	numérico	Número de "me gusta"(opcional)
replies	numérico	Número de respuestas (opcional)
shares	numérico	Número de compartidos (opcional)

4. Pipeline sugerido

1. **Preprocesamiento:** Limpieza básica de texto, normalización y agrupamiento por usuario.
2. **Ingeniería de rasgos:**
   - Rasgos de **comportamiento**: número de publicaciones, frecuencia, promedio de interacciones.
   - Rasgos **léxicos**: longitud promedio, diversidad léxica (Type Token Ratio), repetición de frases.
   - Rasgos **semánticos**: embeddings generados con **Sentence-BERT** o **MiniLM**.
3. **Modelo de detección de anomalías:** Utilizar **Isolation Forest** u otro método no supervisado para identificar usuarios con comportamientos o contenidos atípicos.
4. **Visualización e interpretación:** Construir gráficos con **Plotly** para comparar usuarios según sus rasgos. Mostrar un **ranking de sospechosos** y discutir las razones que justificarían la presencia de bots.

5. Desarrollo de la aplicación

La aplicación deberá:

- Estar desarrollada en **Streamlit**.
- Permitir subir un archivo CSV o usar un dataset de ejemplo.
- Mostrar resultados con visualizaciones interactivas.
- Permitir descargar los resultados (CSV con puntaje de anormalidad).
- Incluir una sección final con la justificación del análisis.

6. Entregables

1. Aplicación **Streamlit** funcional.
2. Archivo `requirements.txt` y carpeta `/src` modularizada.
3. Dataset en formato CSV.
4. Informe breve (2–3 páginas) con:
   - Hipótesis de detección.
   - Explicación del pipeline.
   - Evidencias en Top-K sospechosos.
   - Limitaciones éticas y técnicas.

7. Rúbrica de evaluación (100 pts)

1) Diseño del pipeline (20 pts)

- (10) Selección y justificación de señales (comportamiento, léxico, semántica).
- (10) Integración correcta de BERT + modelo de anomalías.

2) Implementación técnica (25 pts)

- (10) App Streamlit funcional end-to-end.
- (10) Código limpio y reproducible.
- (5) Manejo básico de errores.

3) Visualización y análisis (20 pts)

- (10) Gráficos e interpretaciones útiles.
- (10) Discusión de parámetros y comportamiento.

4) Informe y justificación (25 pts)

- (10) Hipótesis y criterios claros.
- (10) Evidencias y ejemplos analizados.
- (5) Reflexión ética y límites del método.

5) Presentación (10 pts)

- (5) Claridad en la entrega y estructura.
- (5) Calidad visual y argumentativa en la demo.

8. Consideraciones éticas

- No publicar nombres ni identificadores reales de usuarios.
- Los resultados son de carácter **exploratorio**, no determinante.
- Documentar posibles sesgos del modelo o del dataset.
- Promover un uso responsable y educativo del análisis automatizado.

9. Recursos sugeridos

- **Modelo BERT multilingüe:** `sentencetransformers/paraphrase-multilingual-MiniLM-L12`
- **Documentación Streamlit:** https://docs.streamlit.io/
- **Plotly Express:** https://plotly.com/python/plotly-express/
- **Scikit-learn IsolationForest:** https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html

"Detectar patrones inusuales no es condenar; es comprender los límites del comportamiento automatizado."

— Taller Fundamentos de IA