

PRIMERA ENTREGA:  
PREDICCIÓN DE LA VERACIDAD DE PROPUESTAS LABORALES



**UNIVERSIDAD  
DE ANTIOQUIA**  
1 8 0 3

EQUIPO DE TRABAJO  
SAMUEL A. ARISTIZÁBAL GONZÁLEZ  
JULITZA DAZA ZAPATA  
JUAN DAVID LOPEZ AGUIRRE

PROFESOR ENCARGADO:  
RAUL RAMOS POLLAN

UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE INGENIERÍA  
INTELIGENCIA ARTIFICIAL PARA LAS CIENCIAS Y LAS INGENIERÍAS  
MEDELLÍN  
2023

# Predicción de la Veracidad de Propuestas Laborales

Samuel A. Aristizábal González, Julitza Daza Zapata Juan David López Aguirre; Estudiantes U de A

## I. PROBLEMA POR RESOLVER

De un dataset con dieciocho mil descripciones laborales de las cuales ochocientas son falsas se espera realizar un modelo capaz de clasificar las propuestas laborales falsas de las auténticas.

## II. DATASET

El dataset escogido fue tomado de la plataforma Kaggle con el nombre de "[Real / Fake Job Posting Prediction](#)". Las muestras tienen como columnas/variables las siguientes características:

1. **Job\_id**
2. **title** - Título de la publicación
3. **location** - Ubicación del lugar de trabajo
4. **department** - Departamento de trabajo
5. **salary\_range** - Rango de salario
6. **company\_profile** - Perfil de la compañía
7. **description** - Descripción del trabajo
8. **requirements** - Requerimientos para el trabajo
9. **benefits** - Beneficios con el empleo
10. **telecommuting** - 0 o 1 si se realiza teletrabajo
11. **has\_company\_logo** - Si tiene logo de la empresa (0 o 1)
12. **has\_questions** - Si tiene preguntas (0 o 1)
13. **employment\_type** - Tipo de empleo (tiempo completo, medio tiempo, entre otros)
14. **required\_experience** - Experiencia requerida
15. **required\_education** - Educación requerida
16. **industry** - industria en donde se desarrolla el trabajo
17. **function** - Función del trabajo
18. **fraudulent** - Característica de salida (0 o 1)

Con el objetivo de llenar las 30 columnas mínimas pedidas en el enunciado se llenaron 12 variables más con las librerías pandas y random:

19. **country** – País de origen
20. **special\_characters** – (0 o 1) si tiene caracteres especiales
21. **phone\_number** – (0 o 1) si tiene número de contacto
22. **website** – (0 o 1) si tiene pagina web
23. **longevity** – Cuantos años lleva la empresa registrada en la plataforma
24. **vacants** – Numero de vacantes
25. **contract\_type** - Definido o indefinido
26. **comercial\_id** - (0 a 1) Existe un número de registro de la empresa en base de datos gubernamentales o no
27. **multinational** – (0 o 1) Es multinacional o no
28. **number\_employees** – (0 a 1) Informa sobre el número de empleados actuales o no
29. **calification** – (0 a 1) Existen reseñas de los clientes en otras plataformas virtuales o no
30. **deadline** – (0 a 1) Hay fecha límite de la inscripción para la oferta

## III. MÉTRICAS

Algunas de las métricas utilizadas en el laboratorio que se espera utilizaren este proyecto son:

**True Positives (tp):** Esta métrica cuenta el número de casos en los que el modelo predijo correctamente una publicación de trabajo fraudulenta como fraudulenta. Es decir, es el número de casos positivos correctamente clasificados.

**False Positives (fp):** Esta métrica cuenta el número de casos en los que el modelo predijo incorrectamente una publicación de trabajo genuina como fraudulenta. Representa los casos negativos incorrectamente clasificados.

**True Negatives (tn):** Esta métrica cuenta el número de casos en los que el modelo predijo correctamente una publicación de trabajo genuina como genuina. Es el número de casos negativos correctamente clasificados.

**False Negatives (fn):** Esta métrica cuenta el número de casos en los que el modelo predijo incorrectamente una publicación de trabajo fraudulenta como genuina. Representa los casos positivos incorrectamente clasificados.

**Binary Accuracy:** Esta métrica calcula la precisión global del modelo, es decir, la proporción de predicciones correctas en relación con el total de predicciones realizadas

#### IV. DESEMPEÑO ESPERADO

Se espera de este modelo que prediga de manera exitosa con las características de los datos entregados qué descripción de trabajo es fraudulenta o autentica. Identificar cuáles son falsas

ayudaría a prevenir daños económicos a las personas en busca de empleo y la reputación de empresas que pueden ser suplantadas con el objetivo de engañar a más personas.

Un modelo de este tipo puede ser utilizado para aplicaciones como LinkedIn de ofertas laborales para prevenir lo planteado anteriormente

#### V. REFERENCIAS

- [1] Shivam Bansal. "Real / Fake jobs postings" Kaggle, 2020 [Online] Available: <https://www.kaggle.com/datasets/shivamb/real-or-fake-fake-jobposting-prediction>. [Accessed: 26-Sept-2023]