

Primer entrega de proyecto

Por

MARIA ISABEL RUA VELEZ

Materia

Introducción a la ingeniería web

Profesor

Raul Ramos Pollan

**UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE INGENIERÍA
MEDELLÍN 2023**

1. Planteamiento del problema

Para este proyecto se escogió un dataset de propuestas de trabajo, para desarrollar un modelo de predicción que aprenda de las descripciones de trabajo que son fraudulentas. Utilizando un procesamiento de lenguaje natural.

2. Dataset

Vamos a usar el dataset de kaggle **Real / Fake Job Posting Prediction** (<https://www.kaggle.com/datasets/shivamb/real-or-fake-fake-jobposting-prediction>), que tiene 18.000 numero de muestras, con las siguientes características:

- **Job_id**
- **title** - Título de la publicación
- **location** - Ubicación del lugar de trabajo
- **department** - Departamento de trabajo
- **salary_range** - Rango de salario
- **company_profile** - Perfil de la compañía
- **description** - Descripción del trabajo
- **requirements** - Requerimientos para el trabajo
- **benefits** - Beneficios con el empleo
- **telecommuting** - 0 o 1 si se realiza teletrabajo
- **has_company_logo** - Si tiene logo de la empresa (0 o 1)
- **has_questions** - Si tiene preguntas (0 o 1)
- **employment_type** - Tipo de empleo (tiempo completo, medio tiempo, entre otros)
- **required_experience** - Experiencia requerida
- **required education** - Educación requerida
- **industry** - industria en donde se desarrolla el trabajo
- **function** - Función del trabajo
- **fraudulent** - Característica de salida (0 o 1)

3. Métricas

Como métrica de evaluación principal será el error logarítmico medio cuadrático (RMSLE)

4. Desempeño

Lo que se espera de este modelo es una clasificación que con las características de los datos y las metas características prediga qué descripción de trabajo es fraudulenta o real.

También que identifique cuáles palabras de las descripciones son de naturaleza fraudulenta.

5. Bibliografía

- SHIVAM BANSAL - Real / Fake jobs postings | Kaggle, (2020) from <https://www.kaggle.com/datasets/shivamb/real-or-fake-fake-jobposting-prediction>