



MODELO PREDICTIVO DE SALARIOS EN EL SECTOR DE DATOS PARA EUROPA

Autor: Eduardo José Limones Contreras



EL DESAFÍO: ¿CÓMO DEFINIR UN SALARIO COMPETITIVO?

- El mercado de talento en el sector de datos es altamente competitivo.
- Fijar salarios sin datos objetivos genera problemas:
 - Ofertas Bajas*: Pérdida de candidatos valiosos.
 - Ofertas Altas*: Aumento innecesario de costes.
 - Inequidad Salarial*: Desmotivación y fuga de talento interno.

NUESTRO ENFOQUE: DATOS PARA DECISIONES INTELIGENTES

Objetivo de Negocio: Crear un modelo que nos proporcione un benchmark salarial preciso para optimizar la contratación, mejorar la retención y controlar el presupuesto.

Conjunto de Datos:

- Fuente: Kaggle - "Global AI Job Market & Salary Trends 2025".
- Alcance: Filtrado para el mercado europeo, analizando más de 4,500 ofertas de empleo.
- Variables Clave: Nivel de experiencia, ubicación, tamaño de la empresa, habilidades técnicas (Python, SQL, TensorFlow...) y salario.



¿QUÉ GANAMOS CON ESTE MODELO?

- Decisiones Informadas: Pasamos de la intuición a la evidencia. Las ofertas se basan en el valor de mercado real.
- Agilidad en la Contratación: Reducimos el tiempo para definir y aprobar ofertas salariales.
- Transparencia y Equidad: Establecemos una base objetiva para las negociaciones y las revisiones salariales internas.
- Optimización de Presupuesto: Evitamos sobrepagos y asignamos los recursos de forma más estratégica.

EL MODELO EN ACCIÓN: ¿QUIÉN LO USA Y CÓMO?

- **Manager de Contratación (RRHH)**
- *Acción:* Introduce los requisitos de una nueva vacante (Ej: "Data Scientist Senior en Alemania con experiencia en Python y TensorFlow").
- *Resultado:* Obtiene un rango salarial predicho para basar la oferta.
- **Profesional de Datos (Candidato/Empleado)**
- *Acción:* Utiliza la herramienta para comparar una oferta recibida o para preparar una negociación.
- *Resultado:* Entiende su valor de mercado y toma decisiones de carrera informadas.

¿CÓMO LO CONSTRUIMOS?

- **1. Análisis y Preparación de Datos:**
 - Se realizó un Análisis Exploratorio (EDA) para identificar las variables más influyentes.
 - Decisión Clave: Transformamos variables de texto como las "habilidades" en características numéricas que el modelo pudiera entender (One-Hot Encoding).
- **2. Selección y Entrenamiento de Modelos:**
 - Enfoque: Se compararon 3 algoritmos potentes (RandomForest, XGBoost, LightGBM).
 - Optimización: Usamos RandomizedSearchCV para encontrar la mejor configuración de manera automática y eficiente.
- **3. Desafío Principal:**
 - El mayor reto fue convertir la lista de habilidades técnicas (texto libre) en variables numéricas útiles sin perder información de valor.

RESULTADOS DEL MODELO: ¿QUÉ TAN PRECISO ES?

- **Coeficiente de Determinación (R^2): 0.88**
 - ¿Qué significa? Nuestro modelo es capaz de explicar el 88% de la variación en los salarios del mercado. Es un indicador muy alto de su poder predictivo.
- **Error Absoluto Medio (MAE): ~€10,800**
 - ¿Qué significa? En promedio, las predicciones del modelo tienen una desviación de unos 10,800€ respecto al salario real. Es un margen muy aceptable en un sector con salarios tan variados.



INSIGHTS CLAVE: LAS VARIABLES MÁS INFLUYENTES

- El modelo confirma que los factores más importantes para determinar el salario son:
 - Nivel de Experiencia (Senior, Executive, etc.)
 - Años de experiencia específicos
 - Tamaño de la empresa (Grande vs. Pequeña)
 - País de la empresa (Ej: Alemania vs. Irlanda)
 - Habilidades clave como TensorFlow y Python
- Insight de Negocio: No solo importa "ser Senior", sino que el modelo cuantifica el valor adicional de habilidades específicas, permitiendo valorar mejor los perfiles.



CONCLUSIONES

- Hemos construido un modelo robusto y preciso (R^2 de 0.88) que cumple el objetivo de negocio.
- Proporciona una herramienta basada en datos para optimizar la estrategia de compensación.
- Genera insights valiosos sobre los factores que determinan el valor en el mercado de talento.



**MUCHAS
GRACIAS**