

Predicción de Rotación de Empleados usando Modelos de Machine Learning

Resumen Ejecutivo del Proyecto de Ciencia de Datos

Overview

Este proyecto tiene como objetivo construir un sistema predictivo que identifique a los empleados con mayor probabilidad de abandonar la empresa. Se aplicaron técnicas de análisis exploratorio, limpieza de datos, y modelado supervisado para encontrar patrones relevantes que ayuden en la toma de decisiones estratégicas de retención de talento.

Problem

La alta rotación de empleados representa un problema costoso y estratégico para las organizaciones. En este caso, más del 20% del personal histórico ha abandonado la empresa, y no se cuenta con herramientas predictivas que permitan anticipar estos eventos para actuar preventivamente.

Solution

Se desarrolló un modelo de clasificación binaria para predecir si un empleado dejará la empresa (`left_company = 1`). Se entrenaron y evaluaron múltiples algoritmos: Regresión Logística (con y sin pesos ajustados), Árbol de Decisión, Árbol optimizado con GridSearchCV. El modelo final seleccionado fue un **Árbol de Decisión optimizado**, con una precisión del 99% y un recall del 94% para los empleados que se fueron, logrando así una detección temprana con mínimo margen de error.

Details

Datos procesados: 3,290 registros con 10 variables (numéricas y categóricas), Limpieza de valores duplicados, codificación `One-Hot`, y eliminación de outliers mediante el método IQR.

Factores clave detectados:

- Nivel de satisfacción
- Promoción en los últimos 5 años
- Carga de trabajo mensual
- Nivel salarial

Modelos y métricas:

Modelo	Accuracy	Recall (Clase 1)	F1-score (Clase 1)
Reg. Logística (sin pesos)	78%	19%	0.29
Reg. Logística (balanceada)	85%	92%	0.74
Árbol (GridSearchCV)	99%	94%	0.97

Next Steps

- Integrar el modelo en un dashboard para uso por el equipo de RRHH.
- Desarrollar alertas tempranas para empleados en riesgo.
- Ampliar el modelo con datos adicionales (feedback, clima laboral, etc.).
- Realizar pruebas con modelos más avanzados (Random Forest, XGBoost).
- Documentar las recomendaciones de intervención según las variables más influyentes.