

# Proyecto 8:

## Modelo predictivo para la retención de empleados.

---

DAVID FRANCO

| DATA SCIENCE 2024



hack(io)

# Objetivo y Herramientas

- Crear un modelo predictivo que permita identificar el riesgo que tiene un empleado para abandonar la empresa, interpretando los resultados y proponiendo estrategias de retención.
- Hemos utilizado imblearn, scikitlearn, Flask y streamlit.





## Presentación de la empresa

---

Por los puestos, podríamos pensar que se dedica a la manufactura de productos del sector farmacéutico (técnicos de laboratorio, I+D, ventas...), un sector valorado en \$100 billones de dólares en la India.

---

De acuerdo con el conjunto de datos, tendrían 4.410 empleados, aunque el análisis exploratorio sugiere que, sin los duplicados, serían unos 1.500.

---

Sus empleados tienen en 18 y 60 años, algunos de ellos con hasta 40 años en la empresa, lo que sugiere tiene una larga trayectoria.

---

Más de un 60% se dedica a puestos relacionados a I+D, y un 73% tiene formación en carreras de ciencias de la salud.

---

El 60% son hombres. Viven a una media de 10km de la empresa.

## Perfil de los empleados



550

Ingreso mediano mensual en euros (₹49.000)  
En el país, son de media 362€.

2,7

Satisfacción media con el ambiente,  
el trabajo y el equilibrio vida-trabajo.

11

Años promedio en la empresa

37

Edad promedio

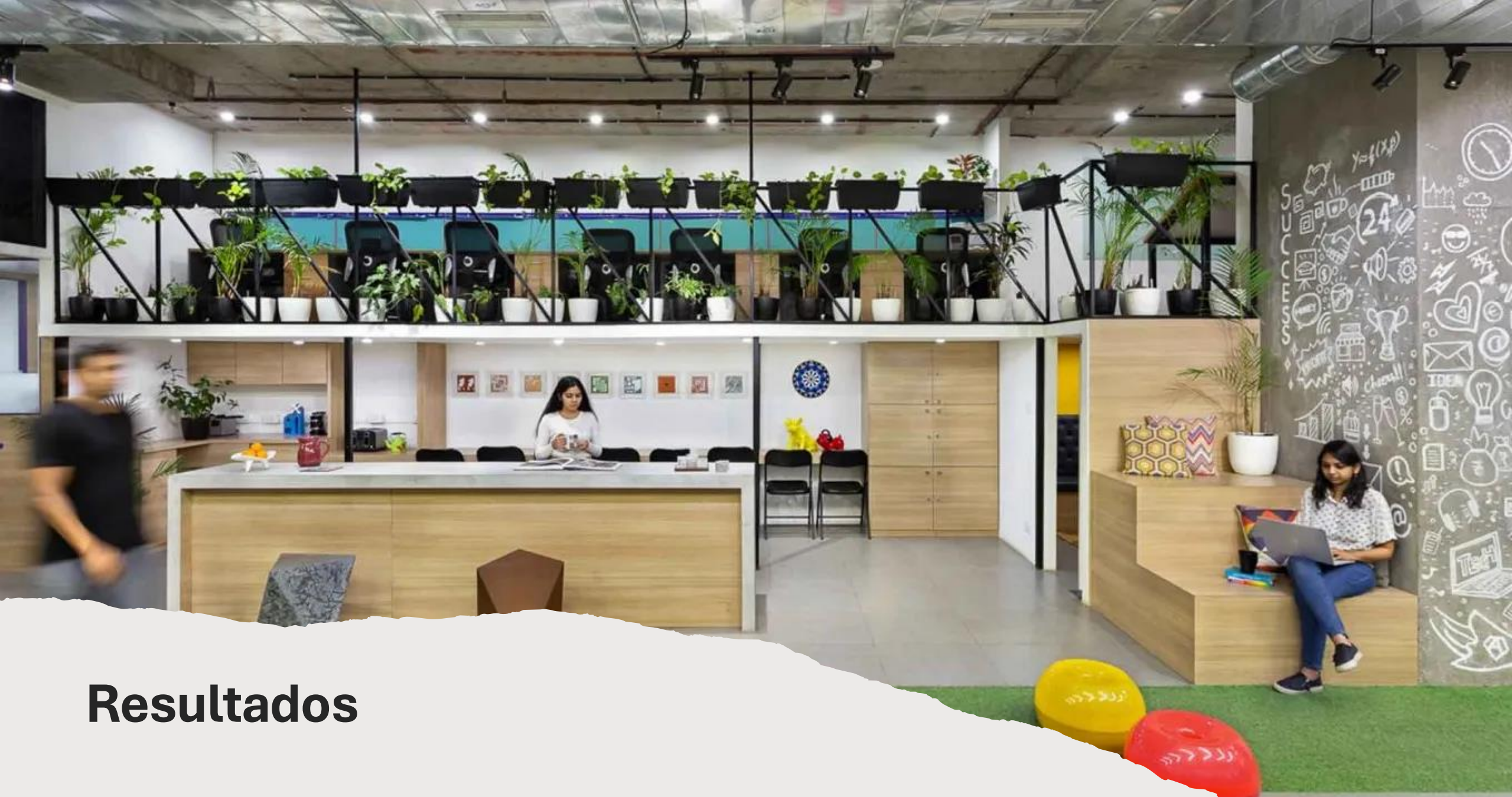




## Por qué la rotación es tan mala para las empresas?

- Cuesta un 33% del suelo anual del empleado saliente.
- Lleva de 8 a 12 semanas sustituir a un trabajador calificado.
- Y entre 1 y 2 meses para que pueda ser productivo.
- *Datos para Estados Unidos de la Society for Human Resource Management (SHRM)*





# Resultados

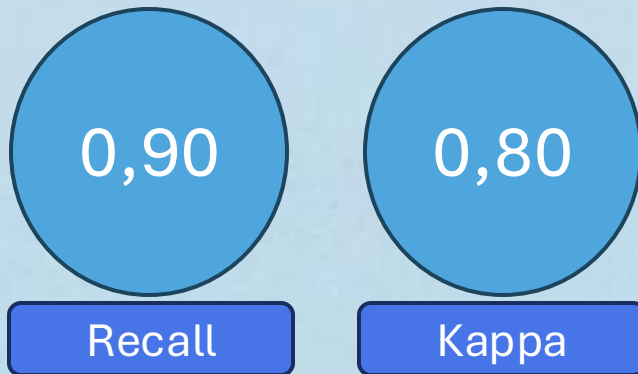
# Resumen de los modelos probados el problema de clasificación

Modelo	Duplicados	Encoding	Outliers	Estandarización	Desbalanceo	Mejor modelo	Recall	Kappa
1	Se mantienen	Catboost	42 eliminados	Standard	No gestionado	XGBoost	1/0.97	1/0.88
2	Se eliminan	Target	16 eliminados	Standard	No gestionado	GradientBoost	1/0.83	0.99/0.24
3	Se eliminan	Target	16 eliminados	Standard	59%-41%	Random Forest	0.95/0.88	0.89/0.75
4	Se eliminan	Catboost	13 eliminados	Standard	59%-41%	XGBoost	1/0.86	1/0.71
5	Se eliminan	Target	15 eliminados	Standard	50-50%	Decision Tree	0.96/0.79	0.93/0.57



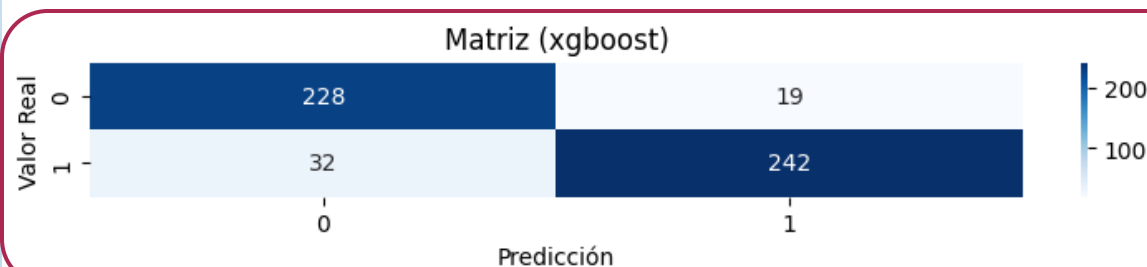
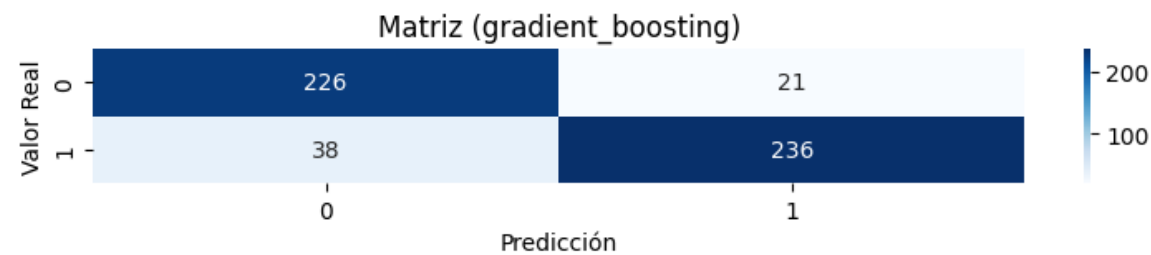
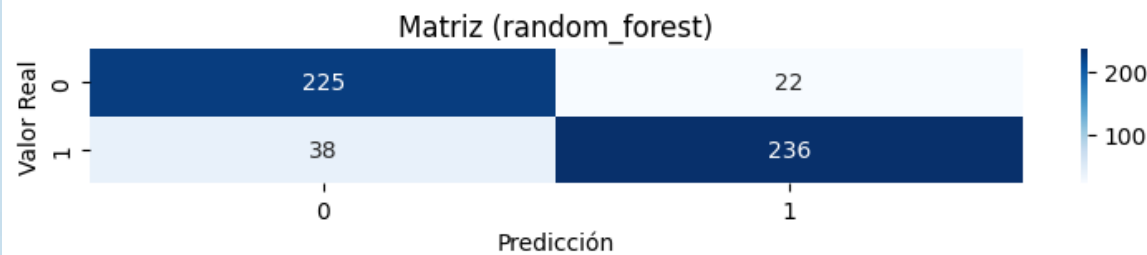
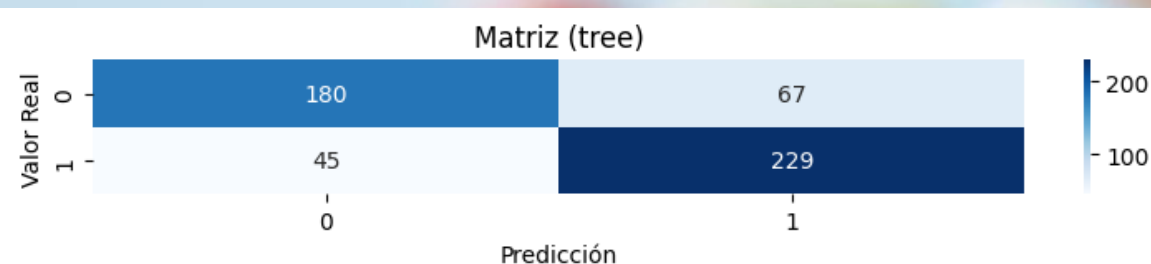
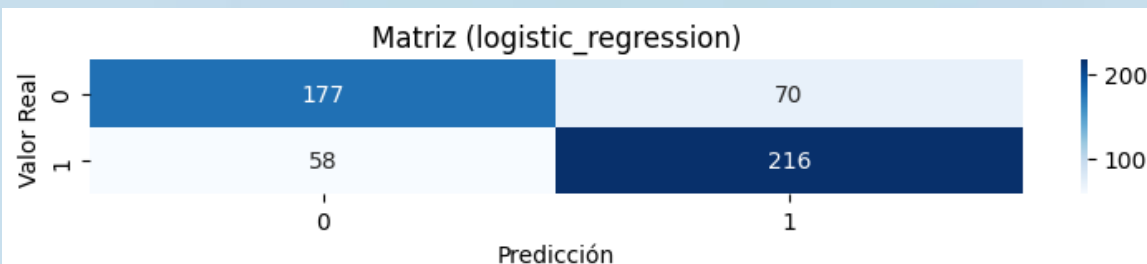
# Decision Tree: dónde preferimos el error.

- Nos interesa priorizar el recall, minimizando los Falsos Negativos.
- Preferimos que el modelo nos diga que un empleado se marchará de la empresa aunque finalmente no sea así, pero queremos evitar el error (Falso Negativo) de decir que se queda cuando finalmente se acabará yendo.





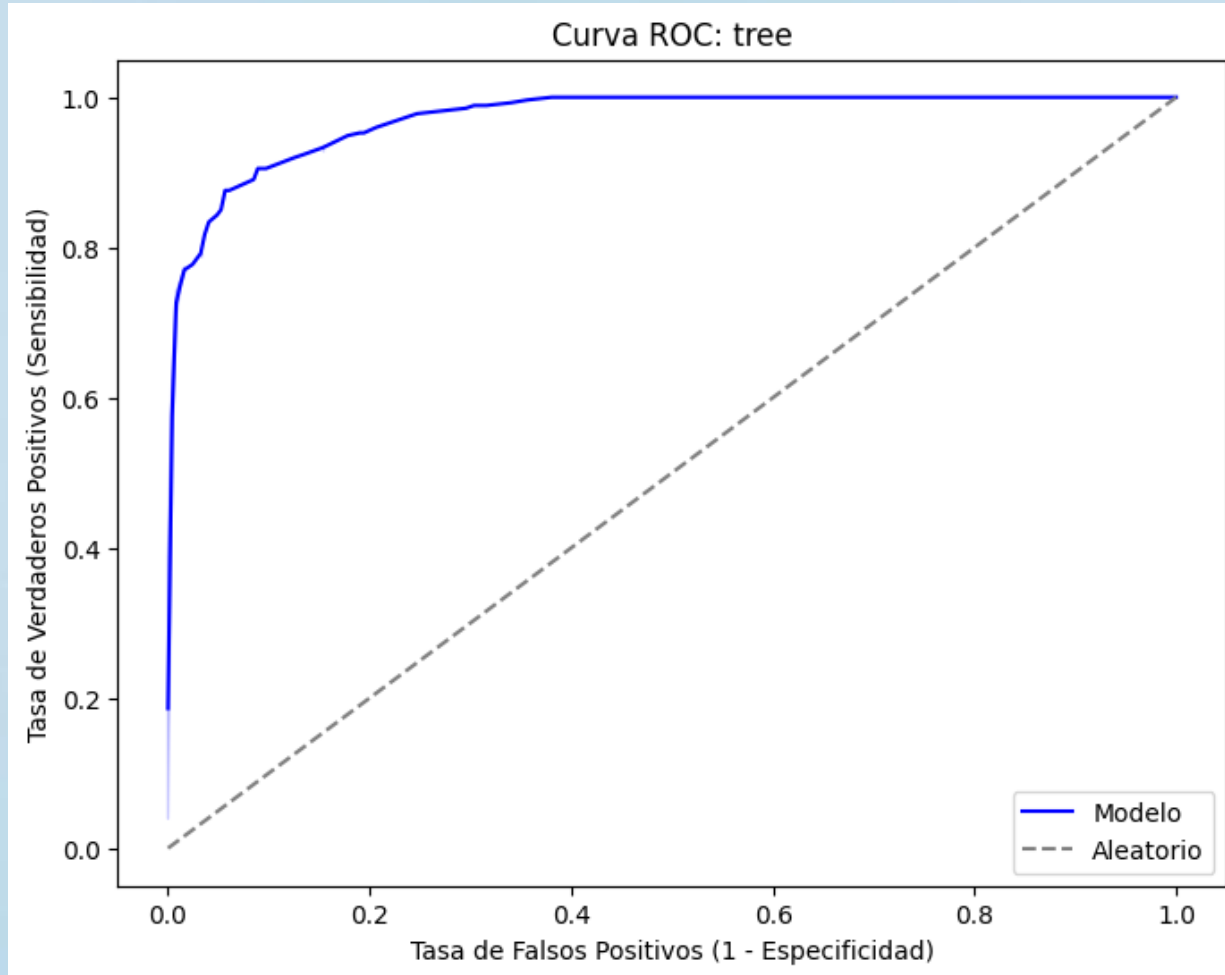
# Matrices de confusión



Nuestro modelo ha identificado correctamente 229 empleados que se van, dejando escapar 45 casos (falsos negativos), lo que representa un 20%.

Aunque otros modelos tienen menor tasa, el overfitting es mayor y podría inducir a otros costes innecesarios para evitar renunciaciones que, realmente, no lo iban a ser.

# Curva ROC



- **Recall alto (FPR):** Detecta correctamente a la mayoría de los empleados que se van.

- **Especificidad (TPR):** Aunque no se prioriza aquí, la baja tasa de falsos positivos sugiere que el modelo también es bueno clasificando correctamente a los empleados que se quedan.

# Gráfica SHAP



## Insights

### Variables importantes

Cuanto más años en la empresa y con el mismo jefe, es más probable quedarse.

Una mayor distancia desde casa y estar soltero aumentan la probabilidad de dimitir.

La satisfacción con el ambiente y el trabajo afectan la probabilidad de quedarse o no.

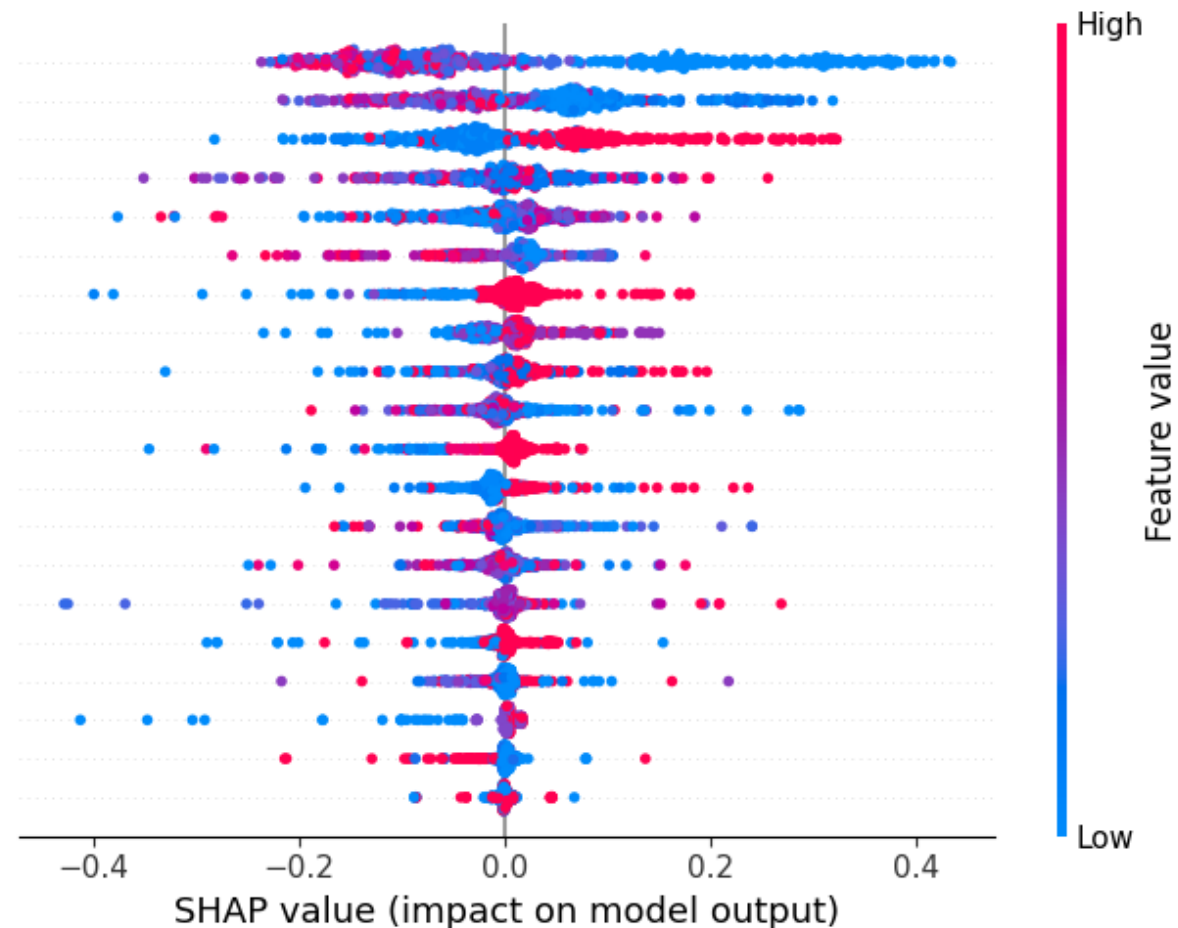
En los sueldos no se muestra una tendencia demasiado clara, aunque se inclina a menores sueldos, mayores probabilidades.

### Menos importantes


La edad, el género, el departamento y el puesto cambian poco la probabilidad.

Las puntuaciones en el trabajo o los años desde la última promoción no parecen tener una tendencia clara.

YearsWithCurrManager\_standard  
YearsAtCompany\_standard  
MaritalStatus  
NumCompaniesWorked\_standard  
DistanceFromHome\_standard  
TotalWorkingYears\_standard  
TrainingTimesLastYear  
JobSatisfaction  
JobLevel  
PercentSalaryHike\_standard  
EducationField  
EnvironmentSatisfaction  
MonthlyIncome\_standard  
Age\_standard  
JobRole  
Department  
YearsSinceLastPromotion\_standard  
BusinessTravel  
JobInvolvement  
Gender

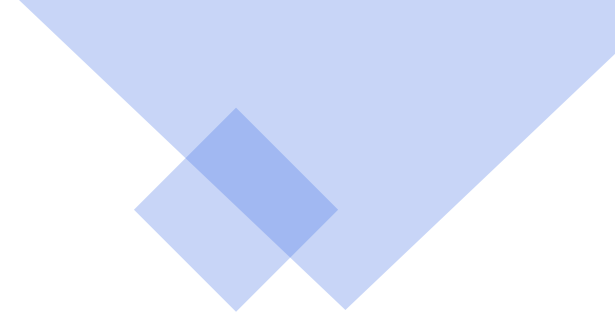
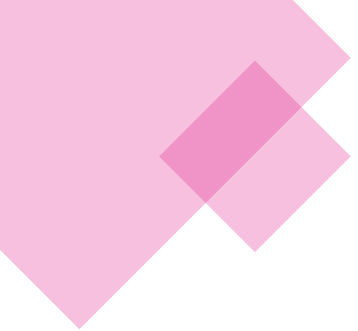






# Recomendaciones para la gestión de los RRHH

- Dar valor a lo “soft”:
  - Los empleados valoran la estabilidad en la empresa y con el mismo jefe.
  - El ambiente de trabajo y la satisfacción con el trabajo son importantes.
- Promover contrataciones de empleados que vivan más cerca.
- Los empleados valoran la formación continua.
- Un mayor sueldo no es garantía de retención.



# hack(io)

