

ANALÍTICA EN RECURSOS HUMANOS LINK AL REPOSITORIO:

https://github.com/EstebanCaroP/proyecto_rhh



1. DISEÑO DE LA SOLUCIÓN

El diseño de la solución propuesta para desarrollar el problema de negocio, se enfoca en la división por departamentos de los empleados. Mediante análisis exploratorio se evidencia que existen tres departamentos en la compañía: Human Resources, Research & Development y Sales.

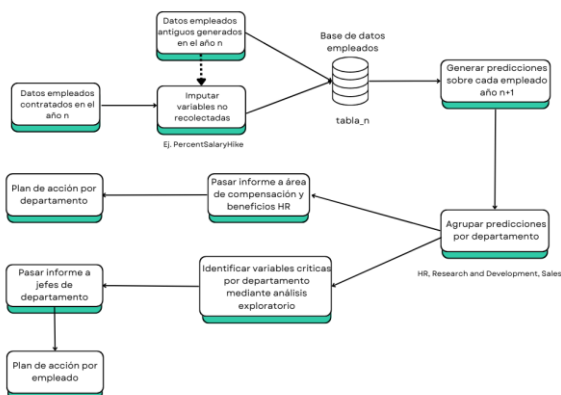


Figura 1. Diseño de la solución

En la figura 1 se observa el flujo del proceso que se debe seguir una vez se tenga el modelo predictivo listo.

2. LIMPIEZA Y TRANSFORMACIÓN

El preprocesamiento del conjunto de datos se realizó mediante consultas en SQL y Python. Se utilizó Python para preparar los tipos de datos que contenían fechas, eliminar columnas que contenían un único valor para todos los registros, y tratar los nulos.

La estrategia de imputación de las variables numéricas y categóricas fue mediante la media para variables numérica y la moda para las variables categóricas. Después de ello se implementan consultas en SQL, en donde se filtran los datos por año para tener la información de los empleados para 2015

separados de la información de los mismos empleados en el 2016. Así, la columna *resignationReason* muestra las personas que renunciaron en el 2016, estas son las etiquetas de la variable objetivo del conjunto de datos del 2015, y de manera similar, nuestro modelo utilizará el conjunto de datos del 2016 para predecir las renunciaciones del 2017.

Finalmente, se unen las tablas que contienen la información general, encuestas de satisfacción laboral, encuestas de desempeño e información de retiro, y se utiliza Pandas para generar las dummies de cada una de las categóricas que fueron seleccionadas dentro del conjunto de datos.

3. ANÁLISIS EXPLORATORIO

El análisis exploratorio se enfocó en el comportamiento de diferentes variables con respecto a la variable objetivo. En la figura 2 se observa la distribución de la variable objetivo, en donde 0 representa a quienes no renuncian, mientras 1 los que sí lo hacen para el año 2016.

Distribución renunciaciones de empleados en 2016

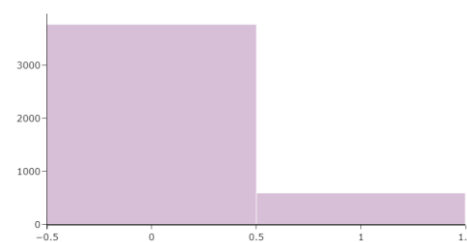


Figura 2. Distribución de la variable objetivo

Hay un desbalance de clases en la variable objetivo, en donde se encuentran 3769 registros de personas que no renuncian, mientras que de quienes sí renuncian se tienen 585 registros. Posteriormente se detallarán las

técnicas para tratar dicho desbalance en el afinamiento de hiperparámetros.

Por otro lado, dado que el diseño de la solución se enfoca en el análisis del comportamiento de la rotación de personal en cada departamento, en la figura 3 se observa el comportamiento de las renunciaciones en cada departamento de la compañía.

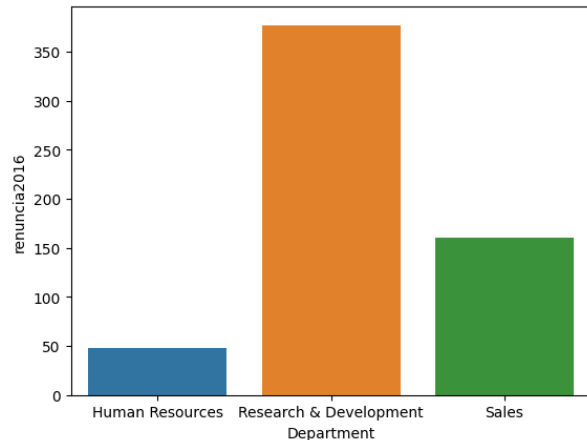


Figura 3. Rotación de personal por departamento

En el repositorio del proyecto se observan otros análisis de interés en el estudio para una comprensión adecuada de los datos.

4. SELECCIÓN DE ALGORITMOS Y TÉCNICAS DE MODELADO

Dado que el objetivo del estudio es realizar un modelo predictivo que permita clasificar si un empleado renunciará o no al año siguiente, se proponen los siguientes algoritmos para la solución del problema analítico:

1. Regresión logística (LR)
2. Clasificador de bosques aleatorios (RF)
3. Descenso de Gradiente Estocástico (SGD)
4. Clasificador XGBoost (XGB)

5. SELECCIÓN DE VARIABLES

La selección de variables se realizó mediante una técnica de eliminación hacia atrás, implementada mediante la librería RFE de

Scikit-learn. Se creó una función que recibiera los modelos descritos en el numeral anterior, y les aplicará a todos la técnica de eliminación hacia atrás. El resultado es un DataFrame que muestra las variables más importantes que cada modelo arrojaba tras la aplicación del RFE, en donde 1 significa una variable de mayor importancia, como se muestra a continuación:

	LR	RFC	SGD	XGB
Age	1.0	1.0	0.0	1.0
Education	9.0	1.0	0.0	0.0
JobLevel	1.0	1.0	0.0	5.0
N°CW	1.0	1.0	3.0	1.0

Tabla 1. Técnica de selección de variables

Se incluyeron las variables que todos los modelos les daba una calificación 1, y con criterios de experto mediante el artículo de Portafolio, se incluyeron aquellas que se consideraban importantes.

6. COMPARACIÓN Y SELECCIÓN DE TÉCNICAS

La métrica de desempeño seleccionada fue el F1 Score, puesto que ésta es ideal para problemas en los cuales hay un desbalanceo de clases.

LR	RFC	SGD	XGB
0.217	0.865	0.253	0.848
0.162	0.866	0.256	0.866
0.227	0.878	0.173	0.829
0.186	0.875	0.090	0.875
0.146	0.867	0.225	0.852

Tabla 2. Métricas F1 Scores para modelos

En la tabla 2 se observa que el Random Forest y XGBoost son los que mejor desempeño tienen, y es el XGBoost el que en promedio tiene mejor desempeño con el conjunto de variables seleccionadas, razón por la cual se elige como modelo final.

7. HIPERPARÁMETROS

La grilla de hiperparámetros que se evaluó se observa en la figura 4.

```
param_grid = [{ 'max_depth': [3,4,5,6],
                 'scale_pos_weight': [6.44],
                 'eta': [0.01, 0.09, 0.1, 0.2],
                 'subsample': [0.5,0.7,0.8,1]}]
```

Figura 4. Grilla de hiperparámetros

maxdepth: controla la profundidad máxima de cada árbol de decisión, valores bajos ayudan a controlar el sobreajuste.

scale_pos_weight: se utiliza para manejar el desbalance de clases. Un enfoque común cuando se trata de clases desequilibradas es establecer el parámetro 'scale_pos_weight' a la proporción de instancias negativas a instancias positivas. En este caso, la proporción 3769/585 es de 6.44.

eta: tasa de aprendizaje, con valores bajos también se puede controlar el sobreajuste del modelo.

subsample: controla la proporción de muestras que se seleccionan aleatoriamente para construir cada árbol en cada iteración. La técnica de afinamiento de hiperparámetros fue la Búsqueda Aleatoria mediante la función RandomizedSearchCV.

	params	mean_test_score
3	{ 'subsample': 1, 'scale_pos_weight': 6.44, 'max_depth': 6, 'eta': 0.2}	0.936284
4	{ 'subsample': 0.8, 'scale_pos_weight': 6.44, 'max_depth': 4, 'eta': 0.1}	0.726852
1	{ 'subsample': 1, 'scale_pos_weight': 6.44, 'max_depth': 3, 'eta': 0.2}	0.658514
0	{ 'subsample': 1, 'scale_pos_weight': 6.44, 'max_depth': 3, 'eta': 0.09}	0.550632
2	{ 'subsample': 0.7, 'scale_pos_weight': 6.44, 'max_depth': 3, 'eta': 0.01}	0.438905

Tabla 3. Métricas del F1 Score para hiperparámetros

Como se observa en la tabla 3, hay un aumento en la métrica de desempeño con respecto al modelo XGBoost al afinar los hiperparámetros.

8. EVALUACIÓN Y ANÁLISIS DEL MODELO

En primer lugar, se analizan las métricas de desempeño en el conjunto de datos de entrenamiento y testeo para el conjunto de datos del 2015. En la tabla 4 se observa que la métrica cae levemente en el conjunto de datos de testeo y nos da indicios de que el modelo

está generalizando bien los datos no vistos y no se está sobre ajustando al conjunto de datos de entrenamiento.

Train Score	Test Score
0.950	0.936
0.948	0.939
0.952	0.924
0.949	0.934
0.949	0.946

Tabla 4. Métricas del F1 Score para train y test

Por otro lado, en la figura 5 se observa que el modelo predijo que 65 empleados renunciarían cuando realmente no lo hicieron. Si bien 65 es un número elevado, no es muy preocupante dado que lo a la larga son personas que no predijo bien, pero que no se van de la empresa, y la mayor pérdida allí serían los costos asociados a la inclusión de dichos empleados en los programas de prevención de deserción laboral, que en últimas tampoco es preocupante porque termina por motivar a dichos empleados que realmente no renunciarán a continuar en la empresa.

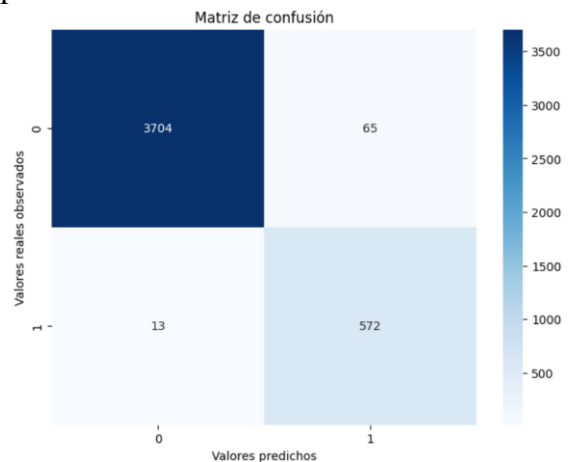


Figura 5. Matriz de confusión

Únicamente 13 empleados se predijeron como falsos negativos, es decir, se predijo que no renunciarían y sí lo hicieron. Realmente es este indicador al que hay que fijar la atención en el momento de evaluación del rendimiento del modelo, y dado que sólo representan

aproximadamente un 3% del total de empleados, se considera que el modelo es adecuado para dar solución al problema de negocio.

9. DESPLIEGUE DEL MODELO

ANÁLISIS GENERAL

En la figura 6 se muestran las predicciones hechas por nuestro modelo para las renunciaciones de cada departamento para el 2017.



Figura 6. Predicción número de renunciaciones por departamento para 2017

ANÁLISIS POR DEPARTAMENTO

Para cada departamento se analizarán las variables que desde la función *feature importances* de XGBoost se evidenció que tienen mayor peso sobre la predicción.

RECURSOS HUMANOS

Calidad de vida

Se agruparon las variables que se relacionan con la calidad de vida del empleado, y su percepción de satisfacción en su puesto de trabajo.

Observamos en la tabla 5 que la diferencia entre quienes renuncian y no lo hacen, se encuentra en la satisfacción con su empleo. Nótese que para quienes renuncian, el valor más común en las encuestas de satisfacción laboral es 2 (Bajo), mientras para quienes se quedan el valor es 4 (Muy alto).

	Enviroment	Job	WorkLife
Renuncian	4.0	2.0	3.0
No renuncian	4.0	4.0	3.0

Tabla 5. Moda de la calidad de vida en RRHH

Se propone identificar a los empleados que tengan una satisfacción laboral menor que 2 y analizar si las funciones que están ejerciendo actualmente están relacionados con el objetivo y el cargo para el cual aplicaron en el momento de vinculación con la empresa.

Edades

Al interior del departamento de recursos humanos, notamos que la edad tiene un impacto en quienes renuncian, en comparación con quienes no lo hacen.

Histograma de edades

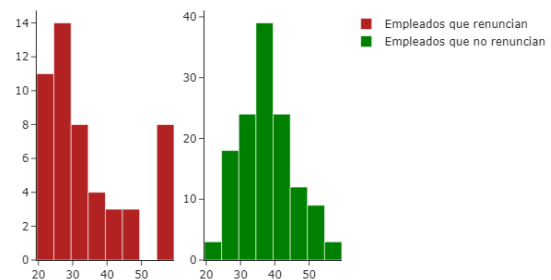


Figura 7. Distribución de las edades empleados RRHH

Nótese en la figura 7 que los empleados que renuncian tienden a ser más jóvenes. Se propone diseñar planes de carrera al interior de la compañía que les garanticen a los empleados una ascenso salarial y estabilidad económica con el paso de los años dentro de la compañía.

Número de años en la compañía

En la figura 8 se observa la variable años en la compañía respecto a las renunciaciones. Nótese que hay un pico para las personas que llevan un año en la compañía, y es cuando más renuncian.

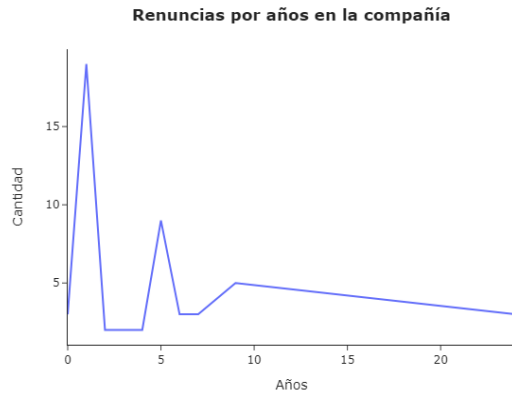


Figura 8. Retiros respecto al número de años en la compañía RRHH

Se propone crear un programa de que recompense a los empleados cuando se cumpla el primer año de trabajo, en el cual puede obtener una retribución económica por su primer año de trabajo o un viaje pago por la empresa, y análogamente cuando lleven cinco años.

INVESTIGACIÓN Y DESARROLLO

Calidad de vida

Para este departamento, el problema radica en la satisfacción con el ambiente laboral como se observa en la tabla 6.

	Enviroment	Job	WorkLife
Renuncian	1.0	3.0	3.0
No renuncian	4.0	4.0	3.0

Tabla 6. Moda de la calidad de vida en R&D

Se propone analizar si las cargas laborales, las relaciones interpersonales y los canales de comunicación se están gestionando de manera eficiente. Realizar programas de prevención del estrés con psicólogos del área de recursos humanos.

Edades

Al igual que en el departamento de RH, se nota que quienes renuncian son en promedio más jóvenes que quienes no lo hacen, dada la asimetría hacia la derecha en quienes renuncian. La estrategia es similar a la de

RRHH, teniendo más posibilidades de ofrecer planes de carrera por la variedad de puestos que hay en R&D.

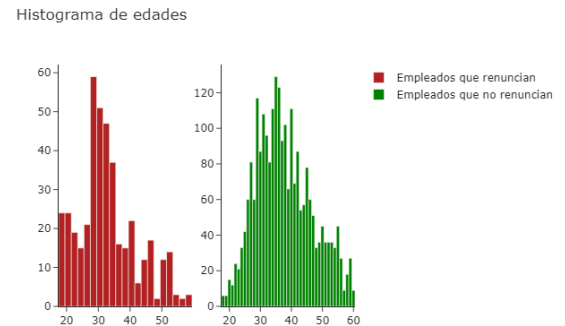


Figura 9. Edades de los empleados en R&D

Número de años en la compañía

Para el departamento de R&D también hay un pico al año de trabajo en la compañía. Se proponen las mismas estrategias que para RRHH, con un cambio en los beneficios a los 10 años de laborar con la empresa.



Figura 10. Retiros respecto al número de años en la compañía R&D

VENTAS

Calidad de vida

De manera similar al departamento de RRHH, los empleados que renuncian en ventas consideran se encuentran en un nivel muy bajo de satisfacción con su empleo.

	Enviroment	Job	WorkLife
Renuncian	3.0	1.0	3.0
No renuncian	3.0	4.0	3.0

Figura 7. Moda de la calidad de vida en Sales

Es necesario realizar una inspección de las cargas laborales y asignación de tareas de los empleados de esta área, dado que los empleados que renuncian puede que no estén encontrando un rol significativo en sus puestos laborales.

Ingresos mensuales

Uno de los principales pensamientos que se suele tener cuando se escucha que alguien renuncia a su trabajo, es que posiblemente no se le paga un buen salario. Veamos que nos dice la figura 11.

Histograma de los salarios mensuales

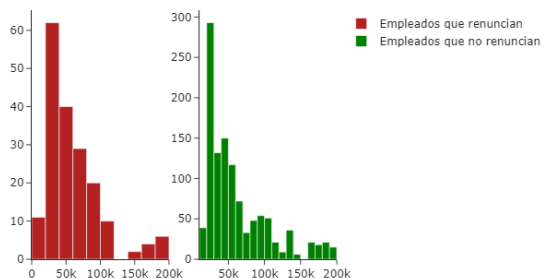


Figura 11. Ingresos mensuales de los empleados y retirados

A pesar de este pensamiento, se puede observar que no existe discrepancia respecto al salario que ganan las personas que renuncian y las que permanecen en esta área. Se analiza esta característica para descartar esta convención en los empleados en la organización.

Número de años en la compañía

La siguiente figura muestra los años en la compañía de los empleados que renuncia.

Renuncias por años en la compañía

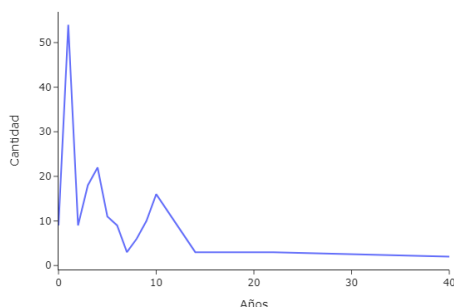


Figura 12. Retiros respecto al número de años en la compañía Sales

Nuevamente se observa comportamientos similares como en los anteriores departamentos, esto es un índice potencial dado que es necesario evaluar qué sucesos se presentan anualmente dentro de la organización. Se propone realizar planes de acción y motivación por periodos anuales, como son las posibilidades de desarrollo personal y consultas sobre satisfacción laboral.

10. CONCLUSIONES

Es necesario tomar medidas preventivas en la rotación de personal dentro de la empresa, principalmente porque la mano de obra dentro de una organización es quien da vida al flujo tanto de procesos como de información. Para el 2017 se tendrá una tasa de rotación de personal del 17,7% por lo que no se alcanzará el porcentaje deseado inicialmente, esto da paso a que la empresa tendrá que verse obligada a tomar todas las medidas sugeridas en cada una de las áreas, con el fin de evitar tanto las renunciaciones como los altos costos que incurren las contrataciones, capacitaciones, curvas de aprendizaje y contrataciones no efectivas.