

INFORME DE CASO DE ESTUDIO DE HR

Valentina Gómez Giraldo
Juan Pablo Ramírez Rincon
Yulianny Alvarez



Facultad de Ingeniería, Universidad de Antioquia
Ingeniería Industrial
Aplicaciones en analítica
Medellín, Marzo de 2024

Enlace proyecto en GitHub:
<https://github.com/YuliAlvarezV/CasoEstudioHR>

Análisis del problema: Problema: La empresa cuenta con niveles de retiros de empleados altos.

Solución: Identificar la causa por la que los empleados se retiran de la empresa para mejorarlas o cambiarlas.

Solución analítica:

- Desarrollar un modelo predictivo para identificar patrones y tendencias en la rotación de empleados.
- Analizar las variables que tienen un mayor impacto en la decisión de los empleados de abandonar la empresa.
- Utilizar estos datos para tomar medidas proactivas y preventivas para reducir la rotación de empleados.

Procesos Impactados:

- Reclutamiento y selección: La alta rotación de empleados aumenta la necesidad de reclutar y seleccionar constantemente nuevos candidatos para llenar las vacantes dejadas por los empleados que se van. Esto puede generar mayores costos y requerir más tiempo y recursos para encontrar y capacitar a nuevos empleados.
- Capacitación y desarrollo: La constante rotación de empleados puede interrumpir el proceso de capacitación y desarrollo de la empresa. Los nuevos empleados

necesitarán tiempo para adaptarse y aprender sus responsabilidades, lo que puede afectar la productividad y la eficiencia del equipo.

- Relaciones laborales y cultura organizacional: La rotación frecuente de empleados puede afectar negativamente las relaciones laborales y la cultura organizacional. La falta de estabilidad en el equipo puede generar incertidumbre y desconfianza entre los empleados restantes, así como una disminución en el compromiso y la moral.
- Retención de conocimiento: Con cada empleado que se va, la empresa corre el riesgo de perder conocimientos y experiencia valiosos. Esto puede afectar la capacidad de la empresa para innovar, resolver problemas y competir en el mercado.

Diseño de solución propuesto: La propuesta de solución al problema de rotación de empleados en la empresa consiste en implementar un modelo analítico para predecir la posible salida de los empleados. Este modelo identificará las variables más influyentes en la deserción y permitirá recomendar acciones específicas al departamento de recursos humanos para la retención de empleados. Además, se evaluarán anualmente los indicadores de retiros anuales y se actualizará la base de datos de los empleados con esta información.

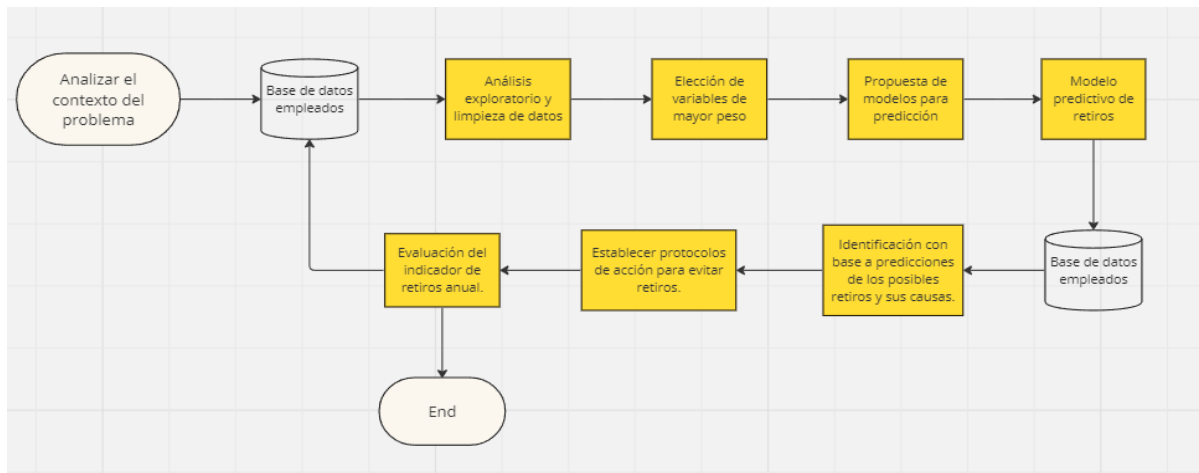


Imagen 1. Diseño de solución

Limpieza y transformación: Para la limpieza y transformación de los datos, se inicia cargando cada una de las bases de datos y se obtiene una visión general de su estructura y los datos presentes para analizar el tipo de datos de cada columna y la cantidad de valores no nulos. Tratamiento de nulos: En la base de datos `employee_survey_data`, se identificaron valores nulos en algunas de las observaciones, por tanto, se reemplazaron estos nulos con ceros, bajo el supuesto de que el empleado se abstuvo de responder la pregunta. Para la base de datos `retirement_info` se identificaron nulos en la columna “`resignationReason`”, esta columna se eliminó ya que no aportaba mayor información. Mientras que para la base de datos `general_data` se encontraron nulos en las columnas `TotalWorkingYears` y `NumCompaniesWorked`, los cuales fueron tratados por imputación y se reemplazaron por la media de cada columna. Transformación de variables: Se cambió el tipo de datos a “object” de las columnas que son cualitativas y para las columnas que representan fechas se cambió el tipo a un formato de fecha. Además, se

crearon diccionarios para renombrar las observaciones de las variables ordinales donde los empleados calificaron algunas características en una escala del 1 al 4, estas

calificaciones fueron reemplazadas por una escala de calificación con adjetivos calificativos, por ejemplo: “Bajo”, “Mediano”, “Alto”, “Muy alto”. Y se eliminaron las columnas innecesarias para el ejercicio, como por ejemplo “`Unnamed: 0`”.

Análisis exploratorio: El análisis exploratorio se hizo mediante la visualización de datos: se hizo un gráfico para cada una de las variables de cada base de datos para visualizar la distribución y el comportamiento de cada variable. A continuación, se muestran algunos de los gráficos.



Gráfico 2. Nivel de satisfacción con el ambiente laboral

Las personas que respondieron que el nivel de satisfacción con el ambiente laboral está entre alto y muy alto representan un poco más del

60% del total, mientras que las persona que dicen tener un nivel bajo y mediano representan aproximadamente un 39%.

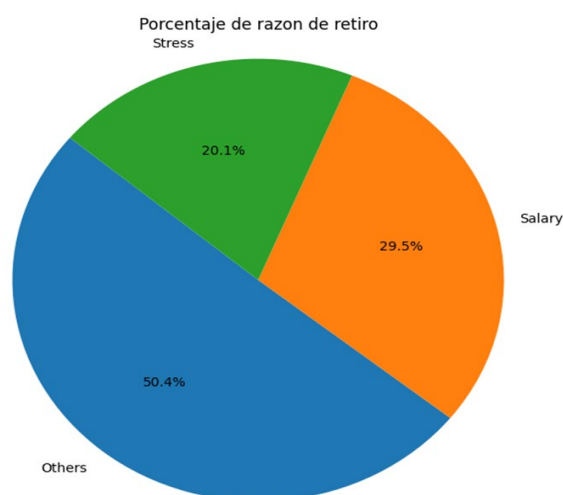


Gráfico 3. Porcentaje de razón de retiro de los empleados.

De los empleados retirados, se puede visualizar que el 50,4% de los retirados no tienen registrada una razón clara de retiro y está nombrada como “Others”. El salario representa un 29,5% de las razones de retiros y el estrés un 20,1%.

Selección de algoritmos y técnicas de modelado: Se seleccionaron 4 algoritmos de modelado teniendo en cuenta el tipo de problema que se está abordando, el tamaño de conjunto de datos, la interpretabilidad del modelo y su rendimiento:

1. Regresión logística: La regresión logística es una buena opción debido a su interpretación sencilla, lo que la hace útil para comunicar las predicciones a los gerentes y profesionales de recursos humanos. También por su capacidad para modelar probabilidades, lo que permite una interpretación intuitiva de las predicciones. Por ejemplo, un empleado con una probabilidad alta de renunciar puede ser identificado y se pueden tomar medidas para retenerlo.

Además, brinda la posibilidad de utilizar técnicas de regularización como la penalización L1 (Lasso) o L2 (Ridge) para mejorar la precisión del modelo.

2. Árboles de decisión: Los árboles de decisión pueden identificar los factores más importantes que influyen en la decisión de un empleado de renunciar, lo que permite a la empresa tomar medidas para retener a esos empleados. Además, funcionan bien con conjuntos de datos grandes, lo que los hace adecuados para empresas con grandes cantidades de empleados.
3. GradientBoostingClassifier: Es una opción más robusta pero también más sólida para predecir qué empleados podrían renunciar en una empresa debido a que es conocido por su capacidad para lograr una alta precisión y su capacidad para manejar características complejas, lo cual es muy útil para nuestro caso donde pueden existir múltiples factores interrelacionados que influyen en la decisión de un empleado.
4. Random forest: Este modelo viene siendo el más robusto y preciso de todos. Y como los árboles de decisión proporcionan una medida de la importancia de cada característica en la predicción y también son escalables para grandes conjuntos de datos.

Con respecto a los factores que se tuvieron en cuenta para seleccionar los algoritmos: la regresión logística y los árboles de decisión son más fáciles de interpretar, mientras que GradientBoostingClassifier y RandomForest son más robustos y precisos.

Selección de variables: Para la selección de variables comenzamos aplicando dummies para representar variables categóricas en forma

binaria (0 o 1), luego separamos la variable respuesta y procedimos a escalarlas, es decir ajustamos las características numéricas para que tengan una escala común sin distorsionar sus distribuciones originales. Ya para la selección de variables se importan diferentes modelos de clasificación: LogisticRegression, DecisionTreeClassifier, RandomForestClassifier y GradientBoostingClassifier ya que son los modelos que anteriormente seleccionamos por su fácil interpretabilidad, capacidad de manejar distintos tipos de datos y su precisión. Luego, a partir de una función usamos estos modelos y un $\text{threshold} = 2.5 * \text{mean}$ (2.5 veces la media) este indica el valor que debe superar la importancia de variables para ser seleccionadas. Esta función realiza una selección de variables basada en la importancia de las características según los modelos proporcionados. Se asigna el resultado de la selección de variables a `var_names`.

Comparación y selección de técnicas: Realizamos una comparación de todos los modelos y sus desempeños y de ahí obtuvimos que DecisionTree y RandomForest fueron los que arrojaron mejores resultados como se puede observar en el siguiente boxplot con sus datos a un lado (el sufijo sel hace referencia a los modelos con las variables seleccionadas). RandomForest siempre destacó por su alto desempeño y por su poca variabilidad lo que nos permite saber que será más estable a los cambios de datos por ello decidimos seguir con este modelo.

```
m_lreg      0.163260
m_dtree     0.669464
m_rf        0.892861
m_gbc       0.338552
lreg_sel    0.119931
rtree_sel   0.672606
rf_sel      0.895756
gbc_sel     0.349970
dtype: float64
```

Tabla 1. Importancia media de los modelos

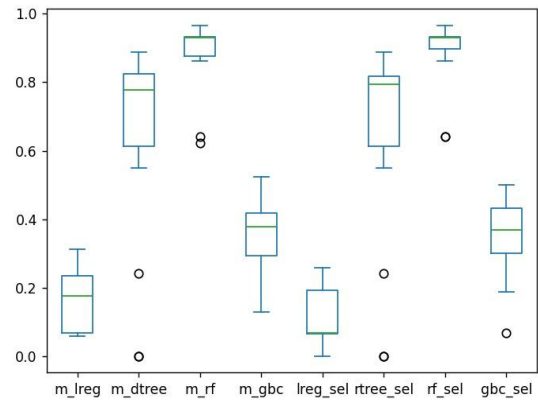


Gráfico 3. Boxplot: desempeño de modelos

Afinamiento de hiperparámetros: Para RandomForest, estimamos el número de árboles de decisiones, número de variables analizadas para la participación de nodos del árbol y el mínimo de observaciones que pueden quedar en un nodo.

Evaluación y análisis del modelo: El RandomForest se desempeña mucho mejor que el resto de los modelos incluso después del afinamiento de hiperparámetros, se revisó el sobreajuste y realizamos una matriz de confusión donde podemos observar que

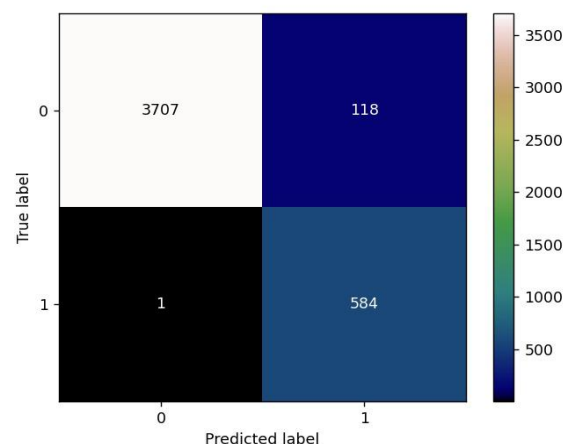


Gráfico 4. Matriz de confusión

Los verdaderos negativos fueron mayores que los falsos negativos que serían los más peligrosos en nuestro caso pues se podrían dejar pasar por alto empleados que realmente quieran renunciar por ello consideramos que fue un

buen rendimiento Finalmente un solo falso positivo realmente nos ayudan a prevenir que en un futuro ese empleado pueda de pronto empezar a pensar en renunciar, pero es un número irrisorio que no afectaron los planes de acción.

Despliegue del modelo: El producto final será un modelo automatizado que podrá ser ejecutado en un periodo de trabajo de un año. Además, se integra el modelo con los sistemas de información y procesos del área de recursos humanos y se les capacita para que continúen suministrando los datos de entrada. Una vez que el modelo esté en producción, se realiza monitoreo y se actualiza constantemente para garantizar que siga siendo efectivo a medida que cambian los datos y demás características de la empresa. También se integra el modelo con los sistemas de información y procesos del área encargada de tomar las decisiones, dicha área tendrá acceso a las salidas del modelo, es decir que recibirá los resultados de sus predicciones de manera fácil y comprensible.

Conclusiones

En conclusión, tras un exhaustivo proceso de análisis y modelado utilizando diversas técnicas de aprendizaje automático, hemos identificado al modelo Random Forest como la mejor opción para predecir la probabilidad de renuncia de empleados en el próximo año. Este modelo ha demostrado un rendimiento superior después de un meticuloso ajuste de hiperparámetros, lo que nos brinda una mayor confianza en su capacidad para proporcionar predicciones precisas y útiles. Al implementar este modelo en un entorno organizacional, podemos anticipar y gestionar de manera más efectiva la posible rotación de personal, permitiendo así tomar medidas proactivas para retener talento clave y garantizar la estabilidad laboral dentro de la empresa.

Recomendaciones

-Análisis de factores de retiro: Identificar y comprender los factores que impulsan la decisión de renuncia de los empleados. Esto podría incluir encuestas de satisfacción laboral, entrevistas de salida y análisis de datos históricos para identificar patrones y tendencias.

-Retención de talento: Implementar estrategias de retención de talento, como programas de desarrollo profesional, oportunidades de crecimiento dentro de la empresa, beneficios competitivos y un entorno de trabajo saludable y equilibrado.

-Revisión de políticas de recursos humanos: Evaluar las políticas de recursos humanos relacionadas con la compensación, el equilibrio entre el trabajo y la vida personal, el reconocimiento y las oportunidades de desarrollo, y realizar ajustes según sea necesario para alinearlas mejor con los objetivos de retención de la empresa.

- Seguimiento y evaluación continua: Establecer métricas y KPIs para monitorear el progreso hacia la meta del 12% de retiros y realizar evaluaciones periódicas para identificar áreas de mejora y ajustar las estrategias según sea necesario.

-Evaluar los perfiles que tienden a quedarse más años para ver que se puede replicar por ejemplo que vivan cerca, ingresos, rangos de edad, cuantos ascensos han tenido, etc.