PRIMER INFORME – RECURSOS HUMANOS

APLICACIONES DE LA ANALITICA 3

Por: Alejandra Aguirre y Aura Luz Moreno

Espacio en GitHub del trabajo de Recursos Humanos: <https://github.com/Moorea-AI/ANALITICA3>

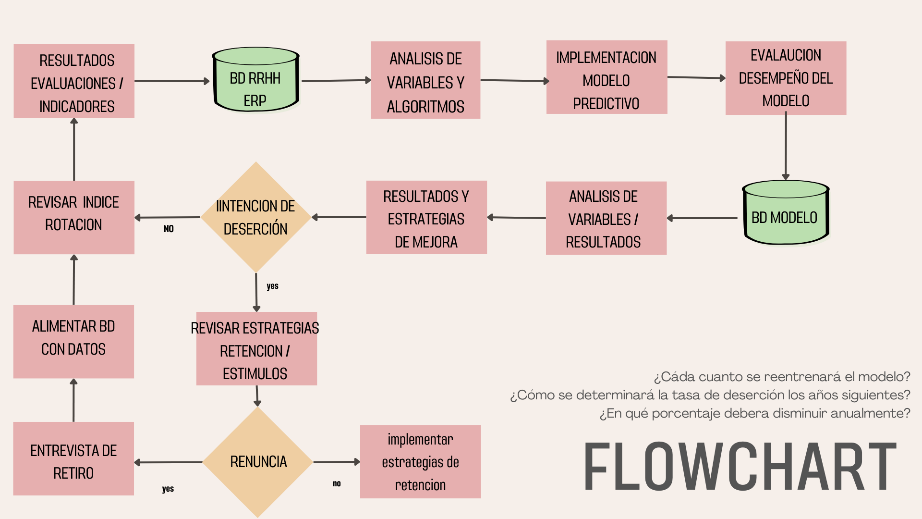
Una empresa que tiene alrededor de 4000 empleados tiene una tasa de retiros de alrededor del 15% anual. Esto quiere decir que el 15% de los empleados en un año se retiran de la compañía por diferentes motivos. El equipo de Analítica debe proponer estrategias que permitan disminuir al menos en un 12% los retiros. A continuación, se presentan el plan de trabajo estratégico para lograr mitigar el problema desde el área de analítica:

1. DISEÑO DE LA SOLUCIÓN

Mediante el uso de algoritmos de analítica, pretendemos predecir que empleado está en riesgo de retirarse, analizando múltiples variables tomadas de la información cotidiana que tiene el área de Recursos Humanos y con base en los resultados poder diseñar planes de prevención de la deserción, reteniendo empleados y procurando un ahorro significativo en costos de capacitación.

Al responder una pregunta tan simple como: El empleado renunciará, solo es posible una variable binaria: Si, No. Con esto entendemos que es un problema de clasificación lo que conlleva a un modelo susceptible de predecir.

Según el artículo Design methodologies for Deep Learning se recomienda realizar diagramas de flujo (Flowcharts) para visualizar la solución del problema, así:



*Figura 1. Flowchart- Elaboración propia*

Según este mismo artículo, podemos seguir unas recomendaciones así:

Encontrar los hechos, los datos, el problema, las ideas, las soluciones y las aceptaciones, entendiéndose estas últimas como lo que es comúnmente aceptado en el área, en la empresa y en el mercado.

Así también, según el artículo “Applying Design Thinking to Artificial Intelligence. Why Should You Use It in Your AI-Based Projects?” existen una serie de pasos para entender mejor el problema entre los cuales esta: Empatizar, Definir, Idear, Prototipar y Testear. Para esto, presentamos la siguiente figura donde detallamos el proceso realizado:



*Figura 2. Design Thinking - Elaboración propia*

1. PREPROCESAMIENTO, LIMPIEZA Y TRANSFORMACIÓN

Las bases de datos son las siguientes:

Las bases de datos entregadas son las siguiente:

**data.dictionay.xlsx:** Descripción de los campos encontrados en las bases de datos.

**employee\_survey\_data.csv**: Encuesta realizada a los empleados sobre satisfacción laboral a final de cada año, se tienen el histórico de dos encuestas, la realizada el **31-12-2015** y la que se realizó el **31-12-2016.**

**general\_data.csv:** Información general de los empleados, dado que la información puede cambiar con el tiempo se extrajo la información con el último corte **31-12-2016** y un histórico con la información con corte **31-12-2015**

**manager\_survey\_data.csv:** Encuesta de desempeño de los empleados realizada por parte de los jefes, se tienen el histórico de dos encuestas, la realizada el **31-12-2015** y la que se realizó el **31-12-2016.**

**retirement\_info.csv:** Información de retiro de los empleados que dejaron la empresa, se tiene la información de retiro de los empleados de 2015 y 2016, sin embargo, la información de 2015 no se pudo obtener completa por problemas en el almacenamiento de la información, por lo tanto, no están todos los empleados que se retiraron ese año.

1. ANALISIS EXPLORATORIO
   1. Revisar la relación entre las encuestas, información general y retiros? Hay algo que decir?
   2. Qué características comunes tienen los empleados que se retiran?
   3. Hay alguna de las áreas que tenga mayor índice de retiro? Porque?
2. SELECCIÓN DE ALGORITMOS Y TÉCNICAS DE MODELADO
   1. Regresión logística: La regresión logística es un algoritmo de clasificación importante, en este caso se podría tener un “se retira”, no se retira, binario. La regresión logística resulta útil para los casos en los que se desea predecir la presencia o ausencia de una característica o resultado según los valores de un conjunto de predictores. Es similar a un modelo de regresión lineal, pero está adaptado para modelos en los que la variable dependiente es dicotómica. <https://aprendeia.com/algoritmo-regresion-logistica-machine-learning-teoria/#:~:text=La%20regresi%C3%B3n%20log%C3%ADstica%20o%20Logistic,%2C%20abierto%20%E2%80%93%20cerrado%2C%20etc>.
   2. Árbol de decisión: Se podría segmentar a los empleados con tasas de retiro altas basadas en características, en este caso es útil el árbol de decisión debido a que el algoritmo de aprendizaje supervisado no paramétrico, se utiliza tanto para tareas de clasificación como de regresión.
   3. Bosque aleatorio: Es una combinación de Arból de decision y regresión logistica, este algoritmo reduce el tiempo dedicado a la gestión de datos y las tareas de preprocesamiento.
   4. SVM o maquinas de soporte vectorial: se utiliza en muchos problemas de clasificación y regresión, como se menciona anterior el modelo es una clasificación binaria, es decir para separar un set de datos en dos categorías o clases diferentes, que es el uso princial de este algoritmo.
3. SELECCIÓN DE VARIABLES
   1. Satisfacción laboral: Un empleado satisfecho probablemente no se retira
   2. Desempeño: Un indicador de desempeño alto puede significar realización profesional
   3. Antigüedad: Una persona con mucho años es poco probable que se retire luego de cierto número de años
   4. Bienestar laboral/oportunidades: Una persona con unos beneficios atractivs dentro de la empresa lo pensasrá antes de retirarse
   5. Salario: Un salario alto, competitivo podría hacer pensar a algunos antes de retirarse.
   6. Clima laboral: Una persona en un clima laboral adecuado con un equipo de trabajo muy bueno también pensará antes de irse.

Y para esto: Matriz de correlación o de confusión?

<https://telefonicatech.com/blog/ml-a-tu-alcance-matriz-confusion>

Podríamos hacer una matriz de confusión para entender que variables están correlacionadas con otras y evaluar el rendimiento del modelo de clasificación.

En la matriz de confusión podemos tener:

Verdaderos positivos: Empleados que efectivamente se retiran y son clasificados como tal

Verdaderos negativos. Empleados que NO se retiran y están bien clasificados

Falsos positivos: Empleados que fueron retirados y que se clasifican como que no

Falsos negativos: Empleados que no fueron retirados y se clasifican como que si

O en una matriz de correlación:

Identificar relaciones lineales entre variables. Correlación con la variable objetivo. Correlación con las variables independientes.

1. COMPARACIÓN Y SELECCIÓN DE TÉCNICAS
2. AFINAMIENTO DE HIPERPARÁMETROS
3. EVALUACIÓN Y ANÁLISIS DEL MODELO
4. DESPLIEGUE DEL MODELO

BIBLIOGRAFIA

Autonomy, H. (2018, 19 junio). Design methodologies for Deep Learning - Humanising Autonomy - Medium. *Medium*. <https://humanisingautonomy.medium.com/design-methodologies-for-deep-learning-fdbc160deff7>

Dorota-Owczarek. (2021, 2 septiembre). Applying Design Thinking to Artificial Intelligence. Why Should You Use It in Your AI-Based Projects? *nexocode*. https://nexocode.com/blog/posts/applying-design-thinking-to-ai/