Módulo 1-Recursos Humanos

Sebastián Buitrago Gómez

Sebastián Ciro Parra

Juan Esteban Marulanda Ayala

Docente:

Juan Camilo España Lopera

Universidad de Antioquia

Facultad de Ingeniería

Ingeniería industrial

Aplicaciones de la Analítica

Medellín

2023-1

a.) Diseño de la solución propuesto

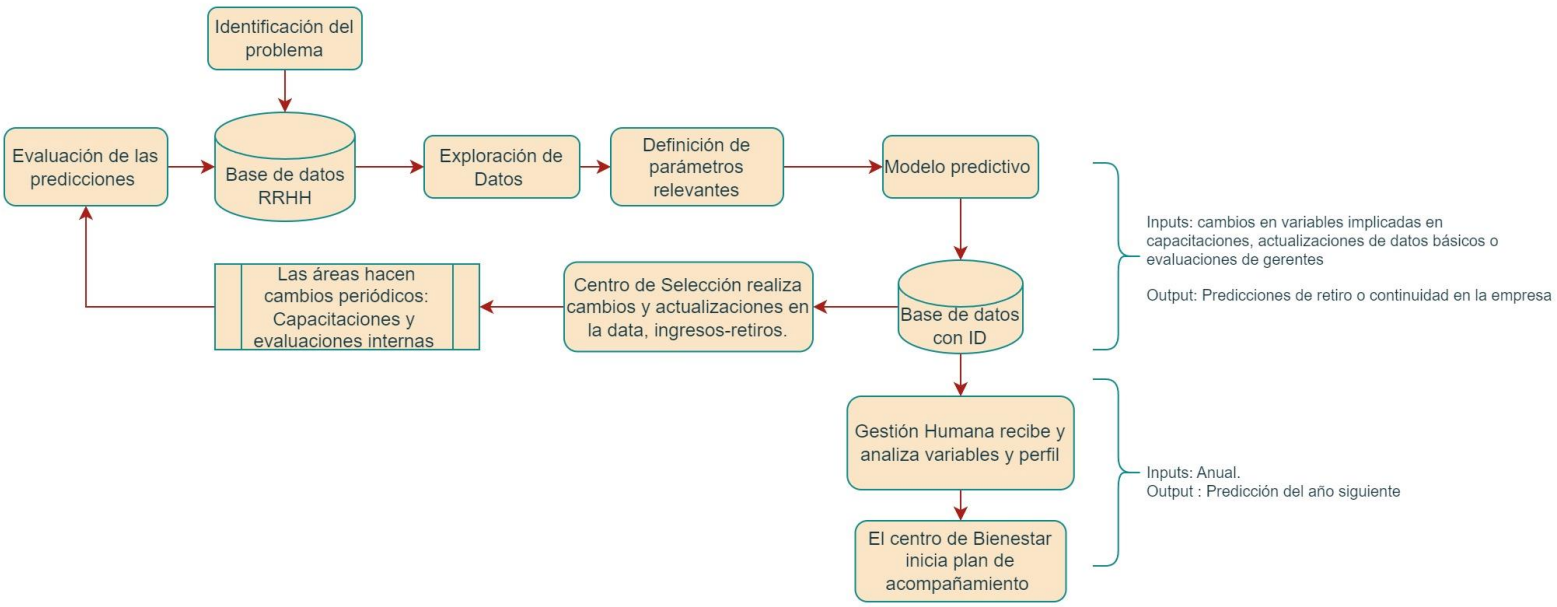
problema de negocio:
En términos generales, el problema es la existencia de una tasa de deserción del 15%, cuestión que según John Badel, gerente general de Lee Hecht Harrison para Colombia, genera un gasto estimado de hasta 12 veces el valor del salario de la persona que se va.

Lo anterior, estimado en los costos del salario de la persona mientras se cubre la vacante –periodo que en promedio puede durar de 2 a 4 meses de aprendizaje de quien asume la posición–, que puede ser de 3 a 6 meses, y del tiempo que la nueva persona toma en alcanzar el desempeño óptimo, calculado en cuatro meses.

Mientras que un análisis de la firma Performia Colombia (proveedor internacional de soluciones para la selección de personal), tomando como base el salario mínimo, señala que las empresas pierden como mínimo 1’155.687 pesos mensuales por contratar mal a un empleado, aunque la cifra puede llegar a ser mucho más alta.

En si, los problemas de las renuncias implican un alto costo para la compañía, análisis de perfiles, selección, entrevistas, capacitaciones y todo lo relacionado con la adaptabilidad del empleado, en sí, se espera entonces mejorar ese novel de renuncias por medio del entendimiento de las variables y de los empleados que están propensos a renunciar, se espera genera un plan de seguimiento y retención que implique el acompañamiento de gestión humana, bienestar, jefes o encargados o incluso de los profesionales o áreas que den lugar

Problema analítico: Definir cuáles son los atributos principales que tienen alguna relevancia en la decisión de retiro de un empleado.
Identificar como es la combinación de las variables que permitirán alimentar el modelo.
Realizar un modelo para predecir si un empleado se va o continua en la empresa. El modelo lo que hará es clasificar un empleado como propenso a irse (si el modelo arroja que se va) o que aun continua (arroja que se queda); es aquí donde se inicia el acompañamiento con gestión humana y el plan de retención.



Limpieza y transformación. Para la limpieza de la información se Realizó una validación de las magnitudes de las bases de datos y de sus consistencias por columna, se realizó una exploración de cuales son los elementos por variables y se buscó encontrar ambigüedades o errores de escritura, previo se analizan los valores nulos y se hace el tratamiento de estos. Se analiza de forma detallada la naturaleza de los datos; en cada una de las bases de datos.

```
#Resumen general
employee_survey.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4410 entries, 0 to 4409
Data columns (total 4 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   EmployeeID             4410 non-null   int64
 1   SatisfaccionAmbiente   4385 non-null   float64
 2   SatisfaccionTrabajo    4390 non-null   float64
 3   BalanceTrabajoVida     4372 non-null   float64
dtypes: float64(3), int64(1)
memory usage: 137.9 KB
```

```
EmployeeID              0
SatisfaccionAmbiente    25
SatisfaccionTrabajo     20
BalanceTrabajoVida      38
dtype: int64
---------
EmployeeID              0.000000
SatisfaccionAmbiente    0.005669
SatisfaccionTrabajo     0.004535
BalanceTrabajoVida      0.008617
dtype: float64
```

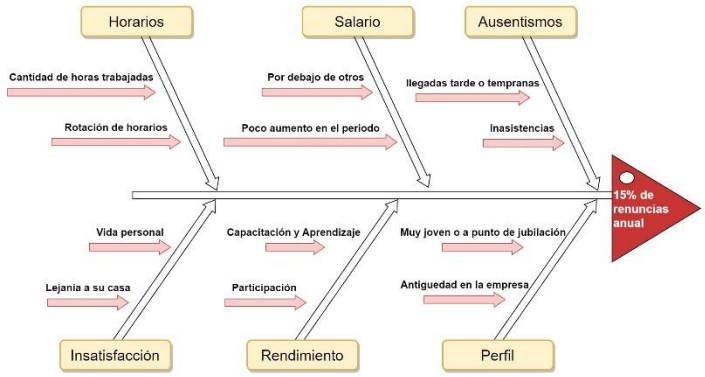
```
3.0    1350
4.0    1334
2.0     856
1.0     845
Name: SatisfaccionAmbiente, dtype: int64
------------------------------
3.0    2660
2.0    1019
4.0     454
1.0     239
Name: BalanceTrabajoVida, dtype: int64
------------------------------
4.0    1367
3.0    1323
1.0     860
2.0     840
Name: SatisfaccionTrabajo, dtype: int64
------------------------------
```

En general, las bases no cuentan con mucha información faltante, hay una variable que por facilidad se transforma a cero, hay 70 datos faltantes en la base de datos de retito, se cambian por otra categoría, se transforma la información de horas de ingresos como la de salidas. El número de compañías trabajadas se autocompleto con un valor.Finalmente se realiza la unión de cada una de las tablas transformadas. Se realizo a partir de la variable ID del empleado. Los datos faltantes a los empleados sin registro de retiro se completan con No Aplica y se utilizó una variable control de retiro y no retiro.

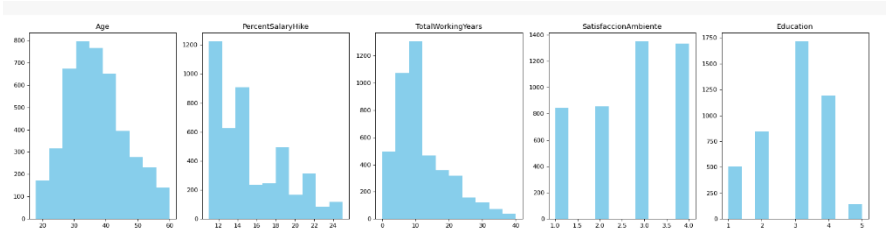
```
[ ]  dt1= employee_survey.merge(general_data, on='EmployeeID', how='left')
     dt2 = dt1.merge(manager_survey_data, on='EmployeeID', how='left')
     dt4 = dt2.merge(retirement_info, on='EmployeeID', how='left')
     dt = dt4.merge(dt3, on='EmployeeID', how='left')
     dt.head()
```

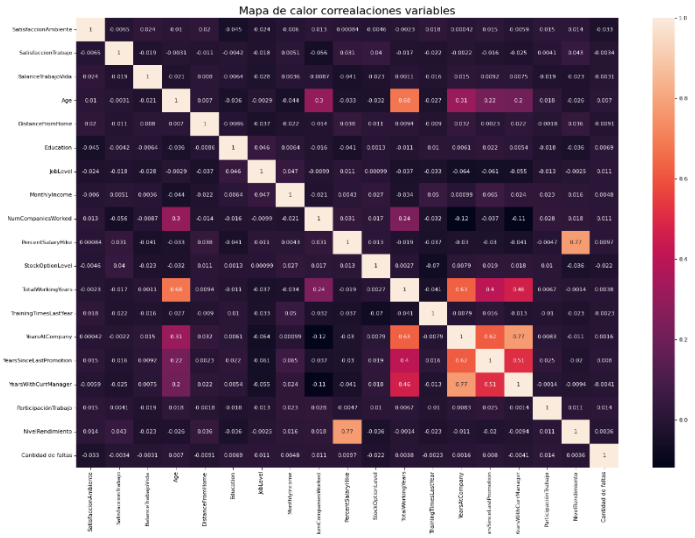
```
Index(['EmployeeID', 'SatisfaccionAmbiente', 'SatisfaccionTrabajo',
       'BalanceTrabajoVida', 'Age', 'BusinessTravel', 'Department',
       'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount',
       'Gender', 'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',
       'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',
       'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',
       'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager',
       'ParticipaciónTrabajo', 'NivelRendimiento', 'Desercion', 'FechaRetiro',
       'TipoRetiro', 'RazonRetiro', 'Cantidad de faltas'],
      dtype='object')
```

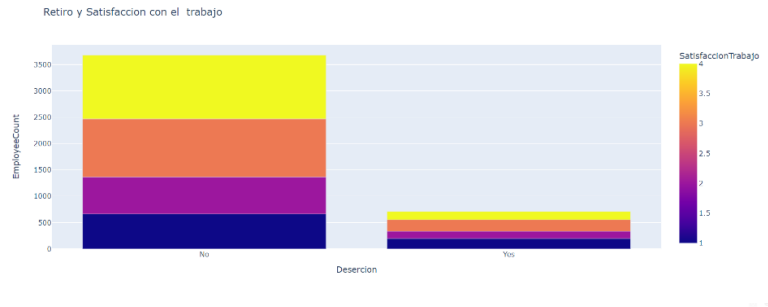
Análisis exploratorio. Dentro del análisis exploratorio se tiene el objetivo de identificar cuáles pueden ser esos focos que representan alguna tendencia a favor del retiro de empleados. Previo al análisis se realizó una valoración de las variables y los atributos relacionados con las posibles razones del problema. Se realizó una espina de pescado para identificar varios focos a la hora de abordar el problema y relacionar los hallazgos de la presente exploración.

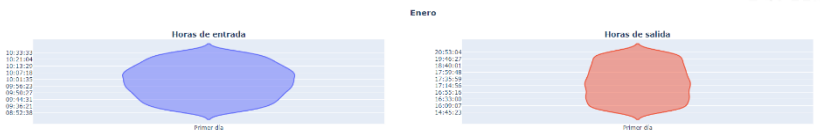


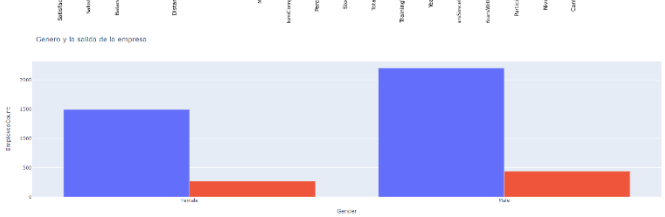
La información obtenida en el análisis exploratorio permite hacer una idea del perfil del empleado, de las características propias en la empresa y las condiciones de renuncia que se presentan, a continuación, se comparten algunos hallazgos de interés en esta etapa.

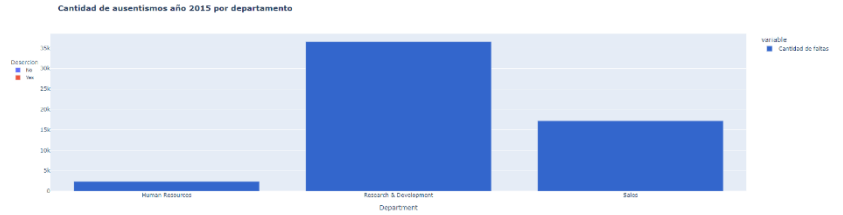




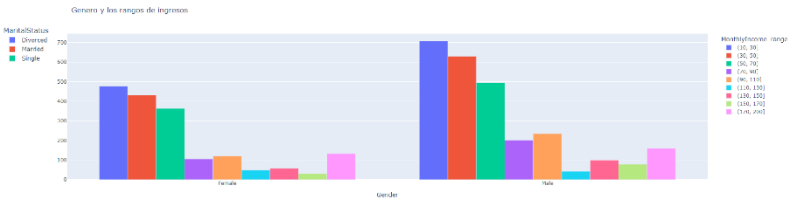


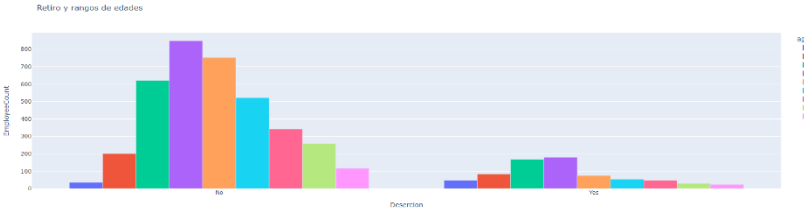


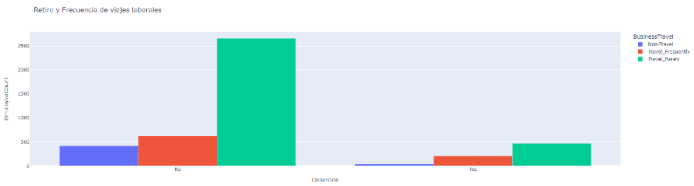




En cuanto a la información referente a retiros se encontró que cerca del 45% de las personas se retiran por razones implícitas a la compañía en las que se puede incluir el plan de acompañamiento. Las edades de renuncias se mantienen entre los rangos de 25 a 35 años para mas del 50% de ellas. Las personas solteras tienden a dejar con mas facilidad su trabajo. Igual que las personas que más viajan por negocios. En cuanto a marcación se tiene que hay un horario de ingreso más estandarizado que el de salida, lo que da a entender que las personas trabajan horas extras regularmente. En cuanto a ausentismos, se tiene que el departamento de i+d+i tiene cerca del 60% al año. Cerca del 45% de las personas que se retiran tienen niveles bajos y medios de satisfacción en el ambiente laboral







La gran mayoría de personas han estudiado en el área de la medicina y ciencias de la vida. las capacitaciones no son un factor de retiro por la uniformidad de los datos y su comportamiento en general. el tiempo que una persona tarda en ascender es relativamente corto. hay personas que viven a más de 25 kilómetros de la compañía

Selección de variables

Se utiliza el escalado y cuantificación de las variables categóricas, se realiza el ajuste de las técnicas que pueden definir una selección de features debida. Selectkbest con F classific y RFE. Se encontraron relaciones entre 8 variables, las cuales lograban definir el modelo entregando resultados coherentes.

Para la selección de los Features se usaron 2 métodos:

- **KBest** que a cada variable le asigna una puntuación y lo que hace es escoger las k variables que tienen el puntaje más alto, en nuestro caso se eligió el valor de k = 8 por lo que se tienen las 8 variables que más peso tienen en el Target.
- El otro método usado es el de **Recursive Feature Elimination (RFE)** que lo que hace es ir eliminando las Features con calificaciones más débiles hasta que se alcanza el número especificado de estos.

Variables elegidas con cada método:

KBest	RFE
SatisfaccionAmbiente	SatisfaccionAmbiente
SatisfaccionTrabajo	SatisfaccionTrabajo
BalanceTrabajoVida	BalanceTrabajoVida
Age	Age
MaritalStatus	MaritalStatus
TotalWorkingYears	TotalWorkingYears
YearsAtCompany	YearsSinceLastPromotion
YearsWithCurrManager	YearsWithCurrManager

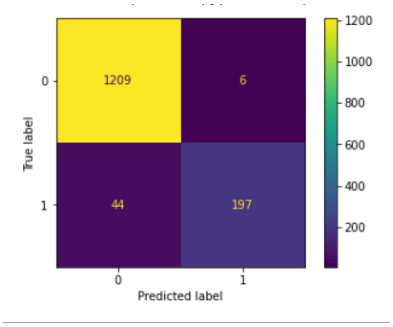
De allí se concluye que ambos métodos tienen 7 de las 8 variables en común.

Comparación de técnicas

	Score		Matriz de confusión		Interpretación
	SVC	KNeighbors	SVC	KNeighbors	
Variables con KBest	0.8390	0.9399	740 | 0 142 | 0	737 | 3 50 | 92	Con las 8 variables que eligió el método KBest como las mejores se tienen que: con el algoritmo SVC se observa que el 83% de las predicciones fueron correctas, mientras que con KNeighbors el 93.99% fueron correctas.
Variables con RFE	0.8390	0.9376	740 | 0 142 | 0	734 | 6 49 | 93	Por otro lado, con las variables elegidas con RFE se obtienen los siguientes resultados: con el algoritmo SVC se observa que el 83% de las predicciones fueron correctas, mientras que con KNeighbors el 93.76% fueron correctas.

Comparación y selección de técnicas

```
KNeighborsClassifier
RandomForestClassifier
ConfusionMatrixDisplay
```



Especificidad, la tasa de verdaderos negativos, ("true negative rate")o TN. Es la proporción entre los casos negativos bien clasificados por el modelo, respecto al total de negativos. 0.96 igual que el promedio de precisión y exactitud

Afinamiento de Hiperparámetros

Con Ramdom Forest Como modelo:

	params	mean_test_score
0	{'weights': 'uniform', 'n_neighbors': 10}	0.168571
1	{'weights': 'distance', 'n_neighbors': 10}	0.982143
2	{'weights': 'uniform', 'n_neighbors': 100}	0.047857
3	{'weights': 'distance', 'n_neighbors': 100}	0.980893
4	{'leaf_size': 10, 'algorithm': 'kd_tree'}	0.964107
5	{'leaf_size': 30, 'algorithm': 'kd_tree'}	0.964107
6	{'leaf_size': 10, 'algorithm': 'auto'}	0.964107
7	{'leaf_size': 30, 'algorithm': 'auto'}	0.964107

Evaluación y análisis del modelo

El modelo asegura que, para el próximo año, cerca de 170 empleados se irán de la compañía, el modelo se ajusta a un periodo anual.

Despliegue del modelo

cargar el modelo para que en el futuro cuando se tengan más datos se puede implementar.