**Credit One**

**Desarrollo de modelos para predecir**

**comportamiento crediticio**

**Elaborado por**

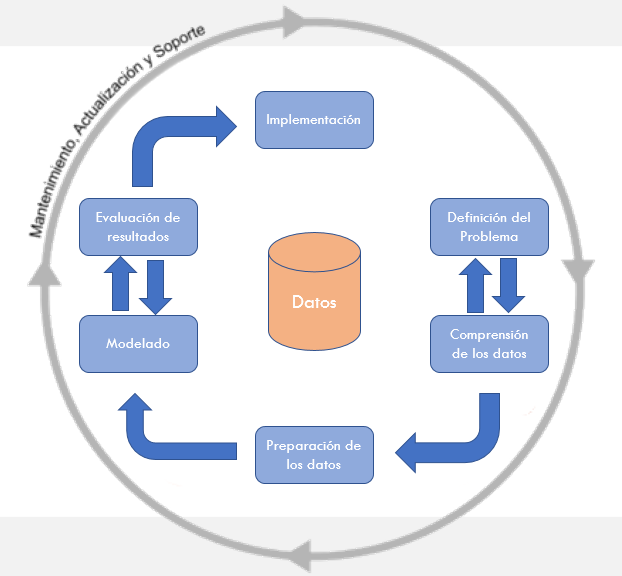
**Michael Fernández Bolaños**

**Descripción y metas del proyecto**

Durante el último año, Credit One ha visto un aumento en el número de clientes que han incumplido los préstamos que han obtenido de varios socios, y Credit One, como su servicio de calificación crediticia, podría arriesgarse a perder negocios si el problema no se resuelve correctamente.

Con base en lo anterior han solicitado la ayuda de nuestro equipo de ciencia de datos para diseñar e implementar una solución creativa y empíricamente sólida que pueda arrojar una predicción sólida sobre el comportamiento de los clientes de Credit One y saber con certeza si cada cliente va a pagar o no el mes siguiente.

**Proceso**

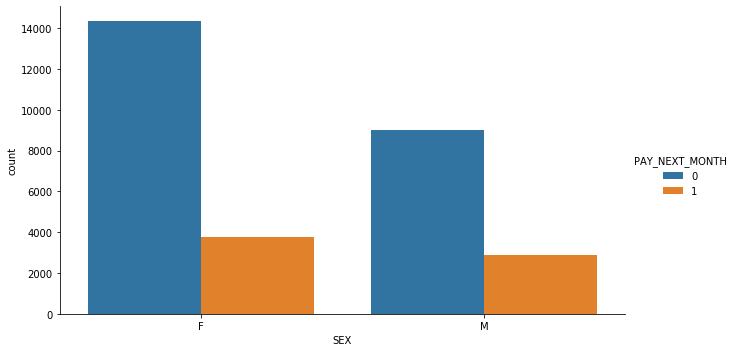
Como aspecto principal en este proceso de ciencia de datos y después de definir el problema los más importantes son comprender y preparar los datos ya que de esto depende poder obtener o no el rendimiento adecuado de los modelos predictivos.

Seguidamente utilizaremos la biblioteca Sci-Kit de Python para la creación y evaluación de al menos 3 modelos predictivos con el fin de evaluar los resultados y determinar cuál se adapta mejor a las necesidades del negocio.

Como paso final se utilizará el modelo para predecir el resultado de nuevos datos de clientes y estos se utilizarán para reentrenar el modelo y obtener mejores resultados con cada ciclo.

**Análisis exploratorio**

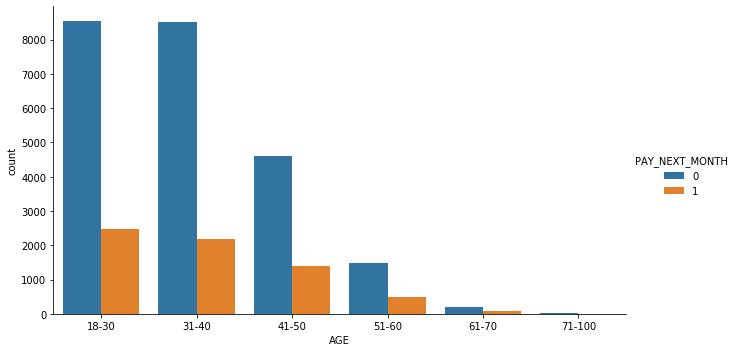
Como parte del análisis exploratorio el primer paso es visualizar la agrupación de las observaciones con respecto a los diferentes atributos, a continuación, podemos ver el comportamiento de pago para el siguiente mes según los diferentes atributos: Sexo, Edad, Educación y Estado Civil



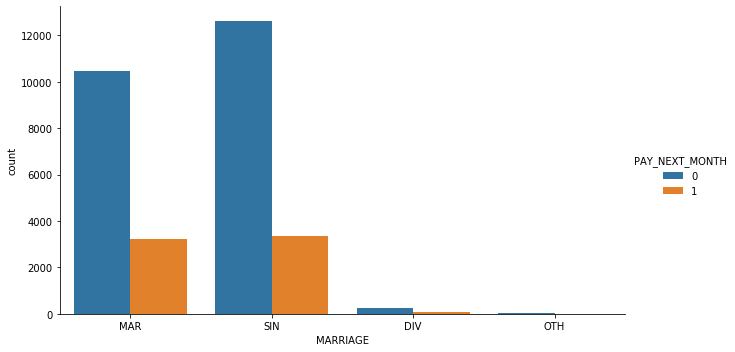
En este gráfico podemos observar como en el caso de los hombres existe una tendencia un poco mayor a realizar el pago del crédito.

Aun así, no parece haber alguna correlación directa entre el sexo y el cumplimiento de las obligaciones crediticias.

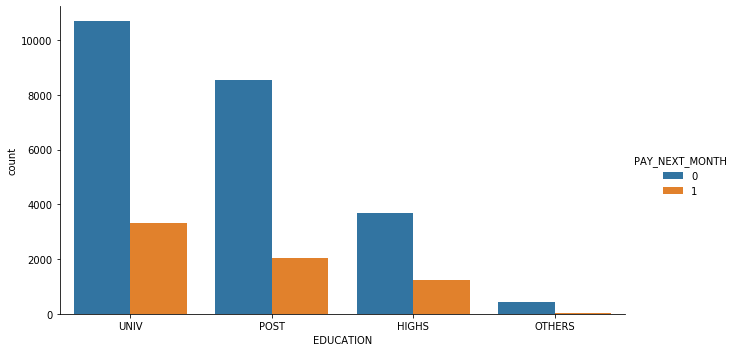
En el siguiente gráfico podemos ver la relación entre el rango de edades de los clientes y el cumplimiento en el pago de los créditos, más sin embargo no se observa una tendencia clara de pago asociado a la edad de los clientes, aunque parece ser que los clientes entre 41 y 60 años tienen un mayor porcentaje de cumplimiento



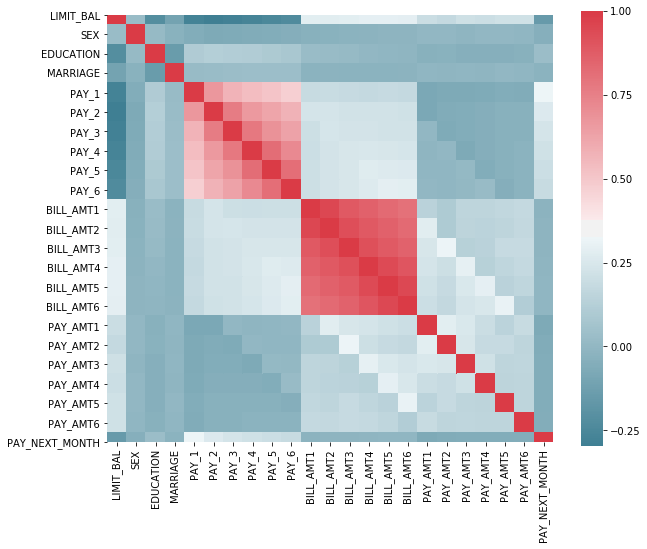
A continuación, podemos observar que al igual que en los casos anteriores, para el atributo de estado civil, no se observa ninguna correlación directa entre el estado civil y las tendencias de pago.



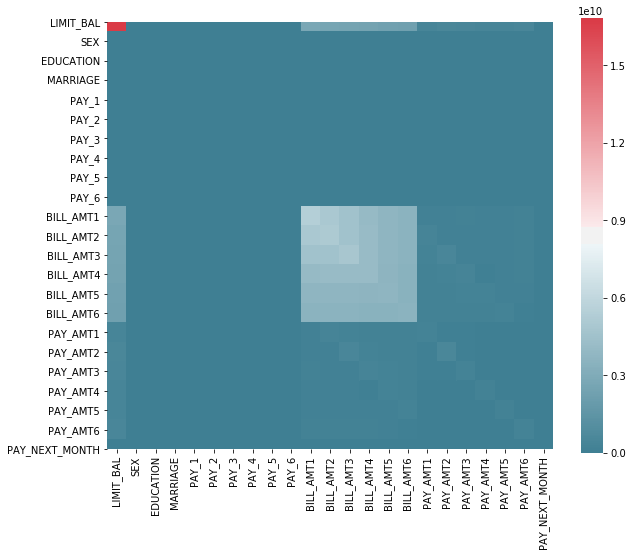
Por último, al analizar la distribución de los diferentes niveles educativos en relación con el cumplimiento en los pagos podemos observar que no se observa ninguna tendencia marcada que sugiera una correlación directa entre el nivel educativo y el comportamiento crediticio de los clientes.



**Matriz de correlación**



**Matriz de covarianza**



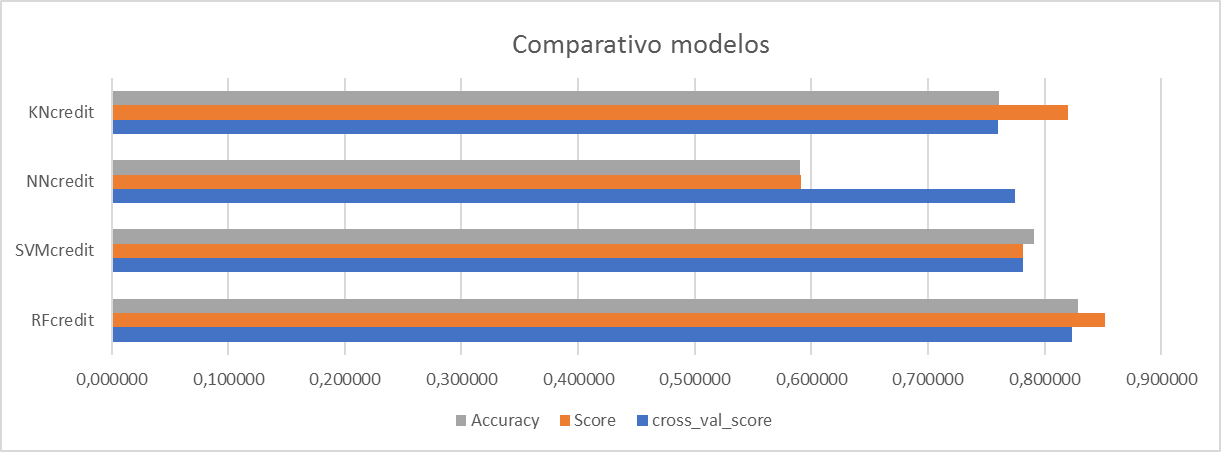
**Métricas**

Con base en la necesidad de análisis se define el requerimiento como categórico y se establece la necesidad de utilizar modelos de clasificación, para ello se utilizaron Random Forest, SVM, Neural Network y KKNN los cuales se configuraron con las siguientes métricas para potenciar las predicciones



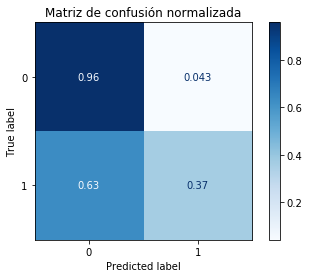
**Modelado**

Con base en la definición de la variable dependiente “PAY\_NEXT\_MONTH” y según el análisis comparativo de los resultados obtenidos podemos notar que el modelo más eficiente fue “FRcredit” con niveles de confianza un poco mayores que los demás modelos entrenados



Como se puede observar en la tabla comparativa las combinaciones arrojaron buenos resultados siendo muy similares las métricas entre ellos, sin embargo, con base en los indicadores se define el modelo FRcredit como la mejor opción para la predicción de los datos.



**Matriz de confusión**

Con base en el detalle de la matriz de confusión podemos notar como según los resultados obtenidos en el modelo elegido, este arroja una mayoría de predicciones como verdaderos negativos lo que se traduce en una correcta identificación de los casos en que el cliente no cumplirá sus responsabilidades crediticias.

Por el contrario, la tendencia respecto a los verdaderos positivos es un poco menos acertada con una mayor tendencia a identificar resultados como falsos negativos lo que implicaría dejar de percibir algunos clientes que en realidad iban a cumplir con sus responsabilidades.

**Conclusiones**

Con base en el análisis del caso de estudio y los resultados obtenidos por el modelo elegido, este funciona como una herramienta de alta certeza para identificar el comportamiento crediticio de los clientes de Credit One.

Claramente no podemos predecir a un 100% cuales clientes van a pagar durante el mes siguiente pero este modelo promete funcionar como una gran herramienta para la toma de decisiones a favor de Credit One.

Según el análisis de los resultados obtenidos por modelo podemos inferir que este se arroja predicciones conservadoras respecto a la calificación de los clientes, lo cual tomando en cuenta los antecedentes de Credit One y tendencias de pago anteriores este parece ser un giro en la dirección correcta.