

Universidad Cenfotec

Data Analytics /Big Data (XTOL)

Entregable # 3

“Build and Evaluate Models”

Estudiante:

Deily Rojas Mejias

Marzo 2020

## Objetivo

Diseñar e implementar un modelo de inteligencia de negocios efectivo por medio de machine learning, que permita aumentar la precisión en el otorgamiento de préstamos y disminuir el índice de clientes morosos, perfeccionando así la plataforma de calificaciones crediticias que Credit One brinda a sus socios. El modelo por desarrollar tomara como base la información de los clientes existentes (P, Crediticia) histórica crediticia de los actuales clientes, permitiendo así generar comportamientos propensos a mora.

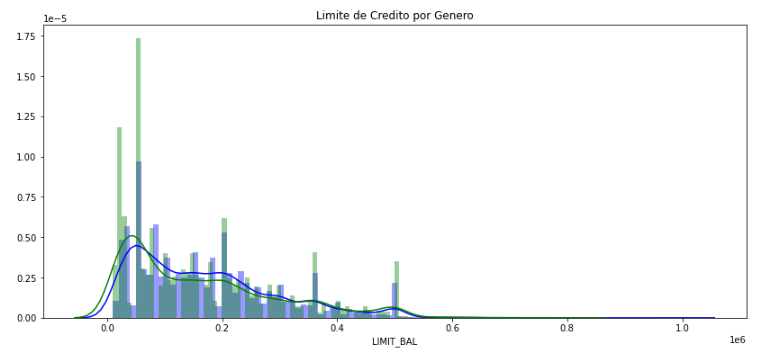
## Administración de la información

Para en análisis del problema presentado en Credit One, se cuenta con un set de información que permite categorizar al cliente según su información personal (Genero, Educación, Estado Civil, entre otros) y su comportamiento crediticio con los diferentes socios de los últimos 6 meses. Por lo cual no se podrán determinar aspectos como eventualidades de los clientes o créditos en épocas de temporadas de consumismo. Por lo cual con la finalidad de perfeccionar el modelo se recomienda realizar los siguientes ajustes:

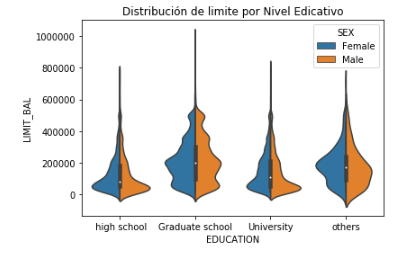
* Aumentar el histórico con la finalidad de aumentar el histórico del crediticio del cliente.
* Incluir datos adicionales de los clientes como lugar de residencia (Provincia, Cantón distrito), Razón de sus ingresos (Asalariado, Empresario) estabilidad laboral (Meses).

Sin embargo, como bien se menciona, la recomendación busca perfeccionar el modelo, con la información recolectada se llevo a cabo un proceso de Ciencia de datos, logrando en este caso aplicarlo en 3 diferentes algoritmos (Random Forest, Support Vector Machine y ADOBoost).

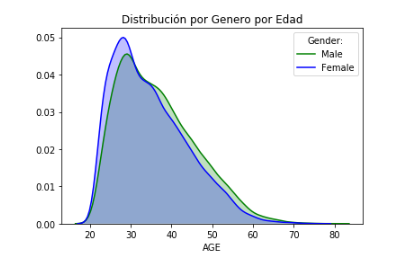
En el análisis de la información se destaca lo siguiente:



Se observan en los datos que es mayor la cantidad de créditos de las mujeres,



Demostrando que la distribución de límite de crédito, por su escolaridad y género.



Finalmente, los créditos por género y su edad son muy similares en el otorgamiento.

## Evaluación del Modelo

Luego de desarrollado el modelo, por métricas de la librería sklearn metrics nos permitirá determinar el score del modelo en su aplicación. Y si es confiable para determinar si el cliente es elegible para el préstamo o no, según sus características. Además, por medio de Matrices de Confusión se evaluará la eficacia del modelo contra el conjunto de prueba para el modelo con mejores métricas.

|  |  |
| --- | --- |
| Modelo | Score |
| Random Forest | 0.79 |
| Support Vector Machine | 0.77 |
| AdaBoost | 0.82 |

Bajo esta perspectiva el rendimiento de AdaBoost es ligeramente superior en su aplicación por lo que será el modelo elegido.

Al aplicar y predecir el set de certificación métricas de efectivas medidas por Rsquared y RMSE

|  |  |  |
| --- | --- | --- |
| Modelo | R Rsquared | RMSE |
| AdaBoost | -0.068 | 0.427 |

Las cuales son métricas positivas, permitiendo con una exactitud del 82% en la predicción de que un cliente se transforme en moroso. Recordando las recomendaciones mencionadas al inicio el modelo no podrá determinar eventualidades que se le presenten a los clientes o comportamientos propios de una temporada por el histórico analizado.