



La Influencia de Python en Data Science para el Sector Financiero

Elaborado por: CARMEN FUERTES BRICEÑO

LIMA 2019

¿Qué es un Programa?

Mano izquierda hacia adelante
Mazo derecha hacia adelante
Mano izquierda hacia atrás
Mano derecha hacia atrás
Mano izquierda a hombro derecho
Mano derecha a hombro izquierdo
Mano izquierda a la nuca
Manta derecha a la nuca
Mano izquierda a caldera derecha
Mano derecha a caldera izquierda
Meneo
Meneo
Salto

Metafóricamente, al lado izquierdo se muestra una serie de instrucciones de música, que deberían ser correctamente seguidas, leámosla y encontremos los 3 errores

¿Qué es un Programa?

Mano izquierda hacia adelante
Mazo derecha hacia adelante
Mano izquierda hacia atrás
Mano derecha hacia atrás
Mano izquierda a hombro derecho
Mano derecha a hombro izquierdo
Mano izquierda a la nuca
Manta derecha a la nuca
Mano izquierda a **caldera** derecha
Mano derecha a **caldera** izquierda
Meneo
Meneo
Salto

Entonces un Programa es: una secuencia de instrucciones que han sido previamente definidas para que se ejecuten correctamente; sin embargo, existen diferentes maneras para que no se cumpla el objetivo, principalmente por errores de sintaxis.

¿Cuál es el Rol del Programador?

Ser el intermediario entre el Hardware/Software y el Usuario Final.

El Programador utiliza el hardware y software como herramientas que le ayudarán con el proceso creativo para solucionar problemas o necesidades del usuario final, a través de una correcta lógica de programación, y es aquí donde elegimos a Python como lenguaje de alto nivel, pero relativamente sencillo de utilizar en comparación a otros lenguajes de programación.

El Origen de Python

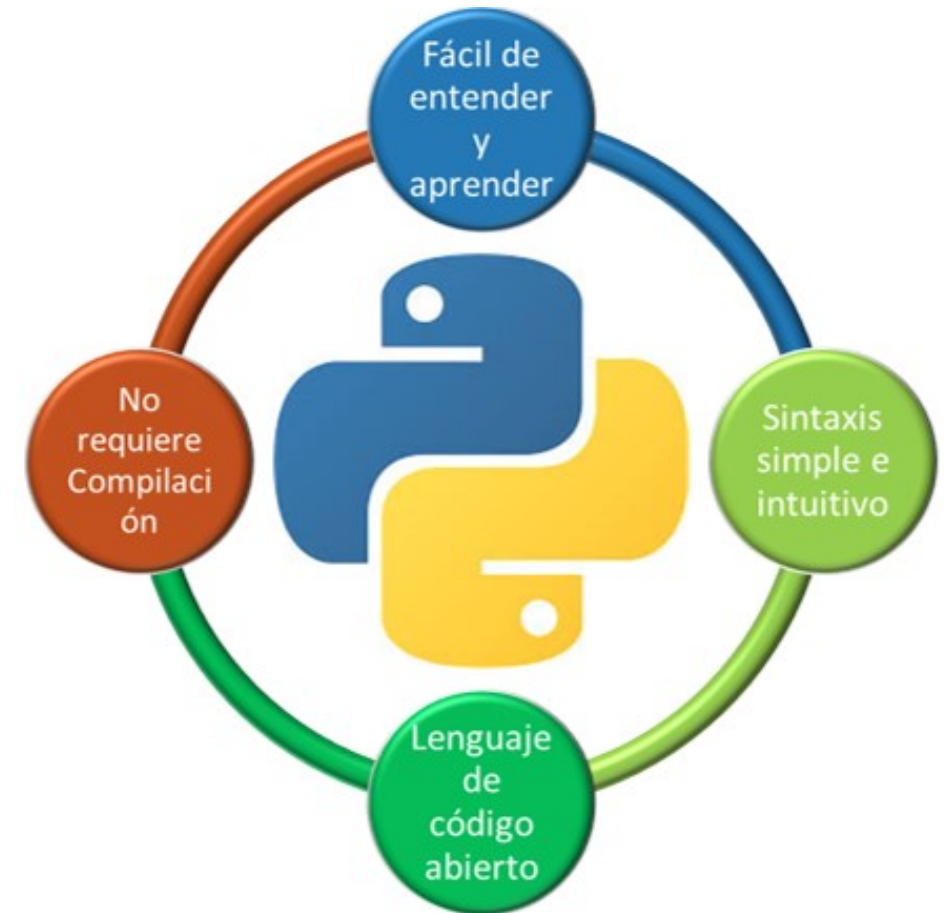
Python fue creado por el holandés Guido van Rossum alrededor de fines de los 80' y principios de los 90'. Nació con el propósito de ser un lenguaje de programación ágil y sencillo, pero considerado hoy en día como un lenguaje de alto nivel por los beneficios de su utilización.



Beneficios de Python

Python es un lenguaje de fácil uso por lo que su principal beneficio es el tiempo que se ahorra en el desarrollo, lo cual nos permite obtener mayor productividad ya que posee una interpretación en Tiempo Real, por lo que no se requiere compilación como en otros lenguajes de programación.

Es uno de los lenguajes que cuenta con una mayor cantidad de repositorios en GitHub.



Estructura de Programa

Pasos Secuenciales

Pasos Condicionales (**If**)

Pasos Repetitivos (**While, For**)

Palabras Reservadas

False	class	return	is	finally
None	if	for	lambda	continue
True	def	from	while	nonlocal
and	del	global	not	with
as	elif	try	or	yield
assert	else	import	pass	
break	except	in	raise	



NO pueden utilizarse como variables

Tips para Sintaxis

---- Utilizado para iniciar comentarios en Python

Int () / float() ---- Conversiones entre cadenas y enteros

Input() ---- Función que regresa una cadena como
entrada del usuario

Las constantes tipos cadena deben utilizar (') o (“)

Las variables pueden iniciar con letras o guión bajo _

Usos Frecuentes

Python es considerado como la herramienta multiuso más popular y de mayor crecimiento en el análisis de datos, ya que es utilizado en muchos sectores principalmente en Finanzas, Data Science, Medicina entre otros, y precisamente por la simplicidad de su lenguaje permite en pocas líneas de código implementar diversas representaciones gráficas y funciones de geolocalización.



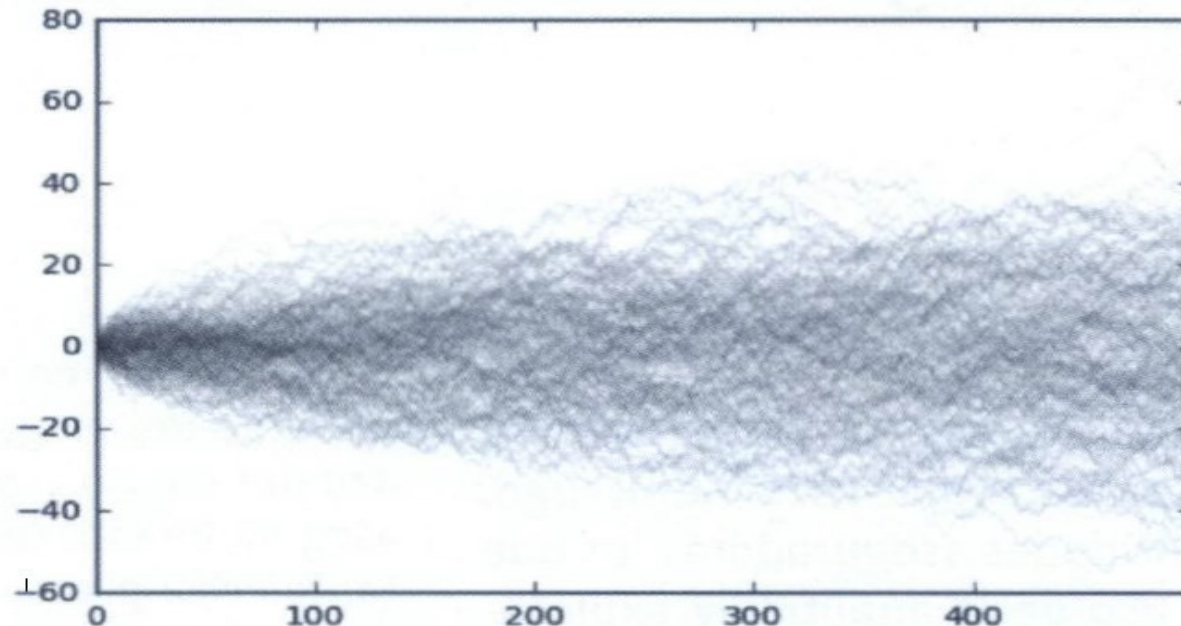
Uso de Librerías Python:

NOMBRE	DESCRIPCIÓN	SITIO WEB
scipy	SciPy (Scientific Python) es una biblioteca de herramientas y algoritmos matemáticos para Python	https://www.scipy.org
numpy	NumPy (Numeric Python) es una extensión de Python, que le agrega mayor soporte para el uso de vectores y matrices.	http://www.numpy.org
pandas	Pandas (Python Data Analysis) es una librería para estructuras de datos, fácil de usar y con herramientas de análisis de datos. Existen librerías complementarias (como PrettyPandas) que ayuda a crear tablas visualmente más atractivas.	http://pandas.pydata.org
matplotlib	Matplotlib es una librería muy completa para la generación de gráficos e histogramas a partir de datos contenidos en listas o arrays.	http://matplotlib.org

Aplicación de NumPy y Pandas: 200 trayectorias como resultado de hacer 500 sucesivos pasos aleatorios

```
import numpy as np
import pandas as pd
datos = pd.DataFrame(np.random.randn(500,200).cumsum(axis=0))
datos.plot(legend=False,alpha=0.05)
```

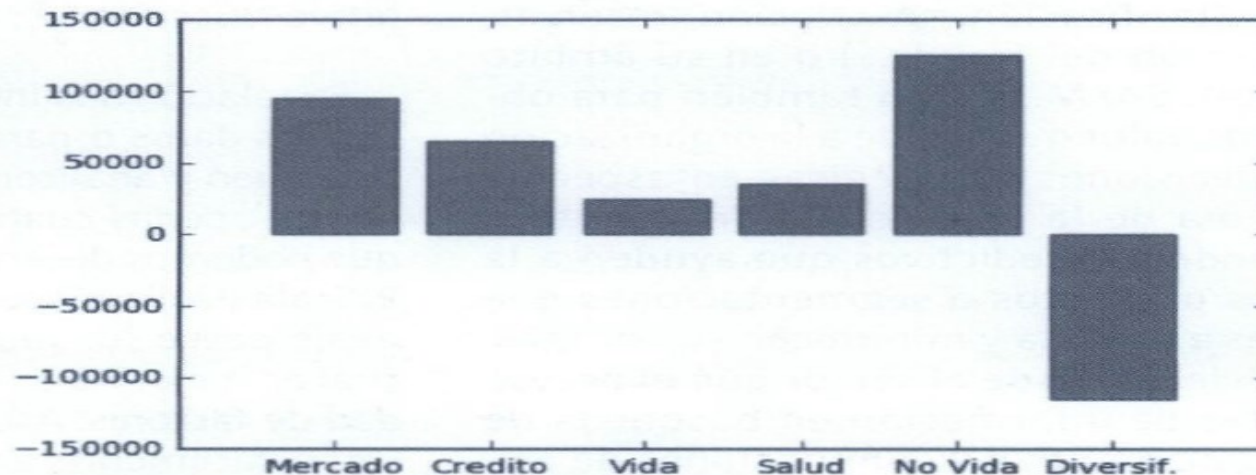
<matplotlib.axes._subplots.AxesSubplot at 0x7f9f21286c50>



Aplicación de Matplotlib y NumPy: Matriz de correlación a diferentes submódulos de Riesgo

```
import matplotlib.pyplot as plt
import numpy as np
SubSCR = np.array([95000.25, 65000.50, 25000.25, 35000.45, 125000.75])
MatCorBSCR = np.array([[1, 0.25, 0.25, 0.25, 0.25], [0.25, 1, 0.25, 0.25, 0.5],
                       [0.25, 0.25, 1, 0.25, 0], [0.25, 0.25, 0.25, 1, 0], [0.25, 0.5, 0, 0, 1]])
BasicSCR = np.sqrt(np.dot(np.dot(SubSCR, MatCorBSCR), SubSCR.transpose()))
BasicSCR_Vector = np.insert(SubSCR, 5, BasicSCR-sum(SubSCR))
print BasicSCR
x = np.arange(6)
plt.bar(x, BasicSCR_Vector, align='center', alpha=0.5)
plt.xticks(x, ('Mercado', 'Credito', 'Vida', 'Salud', 'No Vida', 'Diversif.'))
plt.show()
```

230462.853573



Casos de Uso - Sector Financiero:

Las entidades financieras poseen una gran cantidad de data; sin embargo, existe una gran necesidad de que éstos datos se transformen en información que aporten valor a la organización ya sea para la toma de decisiones en la estrategia de mejorar la experiencia del cliente o en la construcción de modelos predictivos que permitan crear nuevos productos y minimicen cualquier tipo de riesgos.



Credit Report

Your Credit Score is

730

Payment history: EXCELLENT
Amount of debt: EXCELLENT
Credit history duration: EXCELLENT

Applicant Details

<input checked="" type="checkbox"/> Mr <input type="checkbox"/> Mrs <input type="checkbox"/> Miss <input type="checkbox"/> Other			
Surname: DOE	Given Name: JOHN	Middle name: N/A	Phone: 9090 1234 56
Address: 1234 Main St	City: Anywhere	Zip code: 12345	Other Telephone: ()
Work Address: N/A	City: N/A	State:	Mobile Telephone: ()

Scorign Creditício



Objetivo:

Crear un Modelo de Scoring Crediticio a partir de datos extraídos de las redes sociales e integrarlo con un modelo de datos tradicional de préstamos personales y comprobar en qué medida mejora el poder predictivo.

Datos:

Registros: 75,000 Clientes

ROC: 72%

Variables de Clientes adicionales extraídas: 9

Hallazgos:

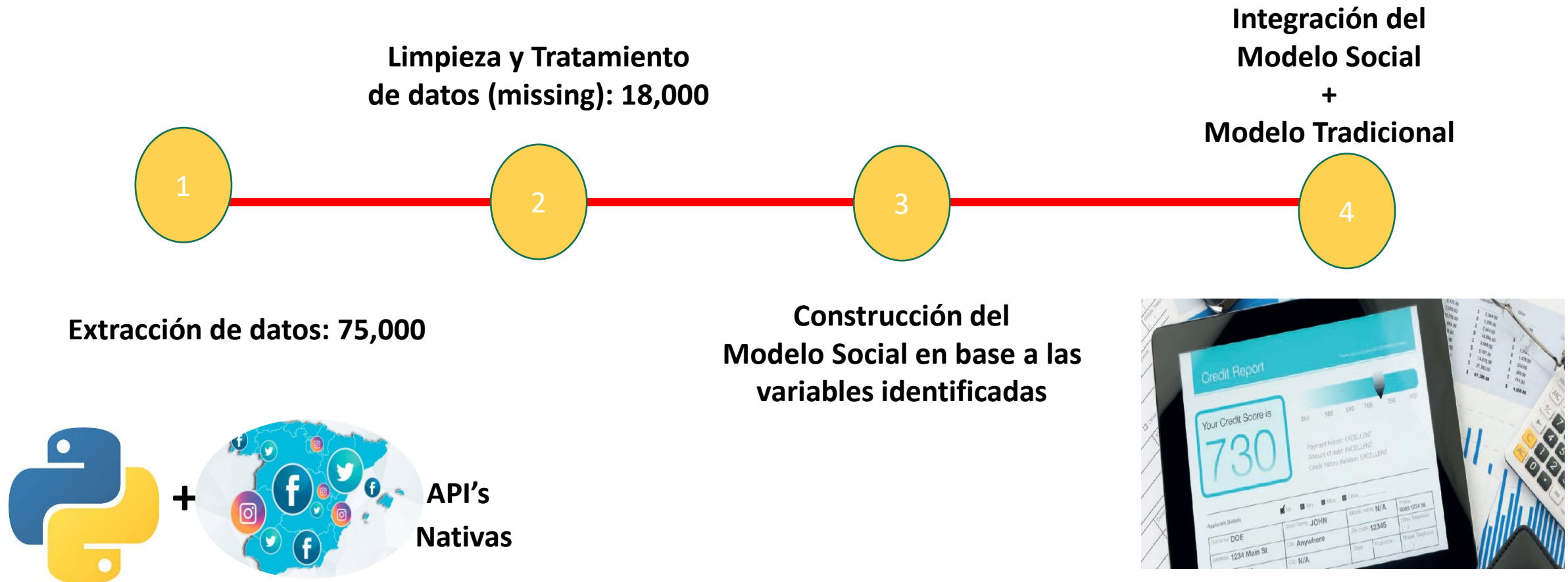
- ❑ La cantidad y calidad de datos de los Clientes en Redes Soiclaes son inferiores en comparación con los datos del banco, ya que solo el 24% tienen datos y solo el 19% tiene información casi completa.
- ❑ Existe una alta probabilidad de error dado que la identificación de cada Cliente no está asociada con su perfil en las redes sociales.
- ❑ La mayoría de variables son cualitativas, las mismas que pueden tomar una gran cantidad de valores que dificultan su tratamiento, pero a cambio enriquecen el tratamiento de la información.

Variables:

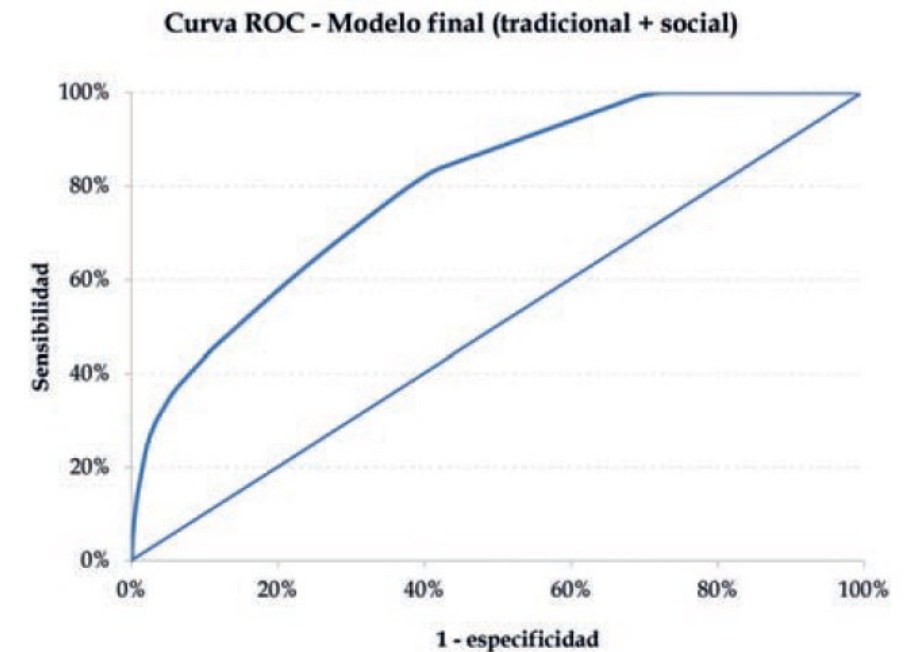
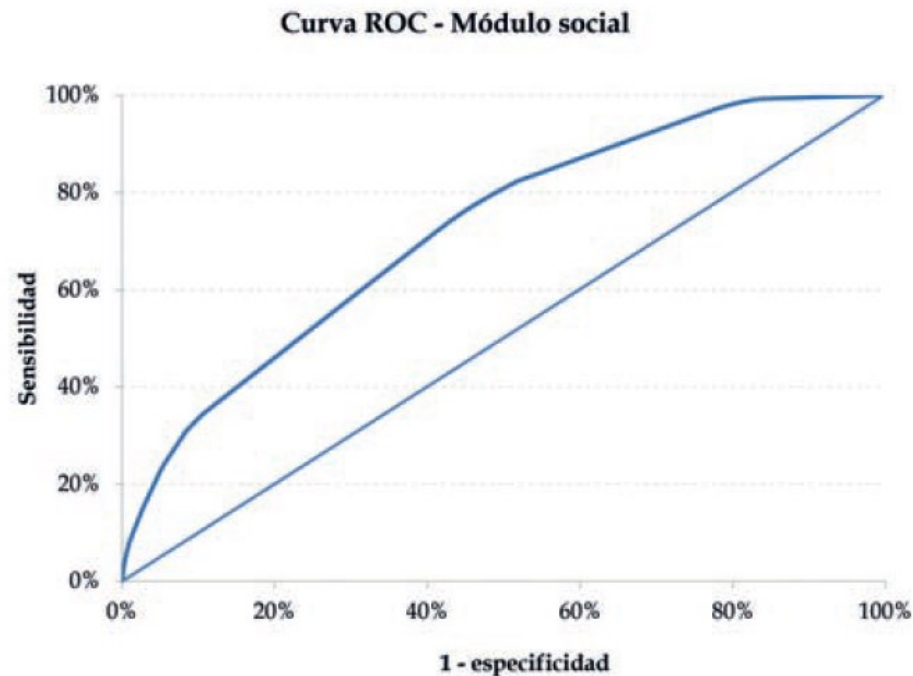
❑ Relacionadas al historial laboral del Cliente (sector, estudios, idioma entre otros).

<i>Variable</i>	<i>Descripción</i>	<i>Peso relativo</i>
Duración cargo actual	Duración en meses del cargo actual	18%
Antigüedad laboral	Antigüedad laboral en meses	15%
Mínima duración en cargo	Mínima duración en un cargo en su trayectoria profesional	15%
Máxima duración en cargo	Máxima duración en un cargo en su trayectoria profesional	13%
Sector de actividad	Categorización INE del sector profesional	12%
Número de trabajos	Número de trabajos actuales e históricos	9%
Tiempo sin estudiar	Tiempo transcurrido desde sus últimos estudios	7%
Idiomas	Número de idiomas que habla	7%
Ratio cargos/años	Número de cargos / número de años de trayectoria profesional	4%

Proceso:



Curva ROC: Mide el Poder Predictivo



Conclusión:

- ❑ El estudio revela que la información de las redes sociales aporta información sustancialmente distinta que complementa y enriquece el modelo tradicional de scoring crediticio. Sin embargo, se evidenció complicaciones en la información a tratar por lo que en el futuro se cree poder conseguir una captura ordenada de los datos para un mayor seguimiento crediticio.

¡Gracias!