

# Proyecto Final – Primera Entrega

Ciencia de Datos Aplicada  
Universidad de los Andes, Bogotá, Colombia

Ivan Saavedra Villamil

Salomón Novoa Montenegro, J. Julián Sanabria Mejía  
{ir.saavedra232, s.novoa485, jj.sanabria}@uniandes.edu.co

Fecha de presentación: septiembre 18 de 2022

## Tabla de contenido

1	Entendimiento del Problema .....	1
2	Ideación.....	2
3	Implicaciones Éticas .....	5
4	Enfoque Analítico .....	6
5	Entendimiento de los Datos.....	6
6	Análisis Exploratorio de los Datos .....	6
7	Conclusiones y Acciones .....	6

## 1 Entendimiento del Problema

Sempli es un *fintech* dedicada a hacer créditos a pequeñas y medianas empresas que estén legalmente constituidas en Colombia. Para este ejercicio, Sempli nos ha suministrado un *dataset* con información del comportamiento de pago de sus clientes. El problema a resolver será clasificar a los clientes que pagarán o no pagarán el crédito de vuelta a Sempli, así mismo identificaremos el perfil de un mal cliente a partir de variables que describen a las empresas incluidas en el *dataset* como el monto desembolsado, la tasa de interés, la cantidad de empleados, el sector económico, etc.

Como empresa de crédito, para Sempli es importante que su estrategia de riesgo crediticio le permita maximizar la utilidad. Es indispensable para Sempli hacer uso del correcto análisis de los datos para aprender cuales son las empresas que tendrán una probabilidad mayor de pagar de vuelta los recursos generando rentabilidad para Sempli y no perdidas en el caso en el que no se pague el crédito.


Para lograr esto, Sempli cuenta con información de centrales de riesgo, sin embargo, esta información no es suficiente ya que usualmente Sempli se concentra en ofrecer la primera oportunidad de crédito formal a empresas con poco historial crediticio. Sempli también cuenta con información propia de clientes a los que previamente le ha desembolsado, los cuales ya

han mostrado un comportamiento de pago durante los más de 4 años que tienen Sempli en funcionamiento.

El objetivo de este proyecto será proveer a Sempli con un modelo de clasificación que le permita identificar de forma más sencilla a los clientes que no pagaran un crédito. La métrica de negocio que se usará para validar el éxito de este ejercicio será el índice de cartera vencida, este índice muestra la proporción del portafolio (dinero prestado a los clientes) que se encuentra con una mora mayor a 30 días sobre el total del dinero que tiene Sempli prestado. El índice de cartera vencida deberá disminuir luego de ejecutar la consultoría a Sempli en comparación con los niveles registrados antes de ejecutar la consultoría.


2 Ideación

Se utiliza *Desig Thinking* para mayor entendimiento del problema a través de los siguientes arquetipos y *journey maps* de las personas involucradas en el problema ya planteado. A continuación, tenemos a Rosa, ella representa a los clientes de Sempli, propietarios de pequeñas empresas que buscan inversión para hacer crecer sus negocios.




Comportamientos y acciones

- Administra su taller de confección.
- Responsable con las obligaciones financieras.
- Apoya y brinda trabajo a madres solteras.
- Disfruta pasar tiempo con sus hijos.



Detalles demográficos

- Madre soltera con dos hijos.
- 42 años.
- Vive en Bogotá.
- Trabaja de forma independiente.



Necesidades y puntos de dolor ('pain points')

- Requiere un préstamo para ampliar su taller y adquirir mejores contratos de confección.
- Las entidades financieras tardan mucho en responder sus solicitudes de crédito y pierde los contratos de confección.




Figura 1. Arquetipo cliente

Rosa es una mujer emprendedora con su propio taller de confección, a través del cual consigue el sustento de su familia y brinda empleo a mujeres cabeza de hogar. El siguiente es su *journey map*.

Actividades	Inicio labores	Visita entidad financiera	Finalización labores
Hace	Contacta a proveedores para cotizar nuevas máquinas para su taller	Lleva sus documentos y los documentos del taller para solicitar un crédito	Revisa los correos del día, no encuentra notificación alguna sobre su solicitud de crédito
Piensa	Si incremento el número de máquinas puedo aceptar contratos de confección más grandes	Ojalá me den el crédito y me lo desembolsen rápido	Esperaba una pronta respuesta, el lunes debo tomar la decisión de aceptar o no un contrato más grande.
Oportunidades	Con mayor número de máquinas, podría contratar más madres solteras	Podría estar haciendo algo más productivo que esperar en la fila del banco	Disminuir el tiempo de respuesta de la entidad financiera

Figura 2. Journey map cliente

La siguiente persona involucrada y muy importante es Alejandra, ella representa a los analistas de riesgo que atienden las solicitudes de créditos realizadas por los clientes de Sempli.



Figura 3. Arquetipo analista de riesgo

Alejandra es una profesional que trabaja para Sempli en el área de análisis de riesgo, siempre busca realizar su trabajo de forma más eficiente y responsable, el siguientes es su *journey map*:

Actividades	Inicio labores	Revisión correo	Reunión
<b>Hace</b>	Digitación de información, análisis de resultados y cálculo de riesgo.	Revisa el reporte de clientes en estado de mora	Recibe queja de clientes por negación de crédito y tiempos de respuesta a las solicitudes
<b>Piensa</b>	Hoy voy a analizar el riesgo de la mayor cantidad de las solicitudes de crédito.	Para algunos de estos clientes, yo aprobé su crédito	Hago mi mayor esfuerzo y aún así no es suficiente
<b>Oportunidades</b>	Evitar el tiempo y esfuerzo de digitar la información del cliente	Mejorar el cálculo del riesgo para los clientes	Aumentar la cantidad de analistas de riesgo y/o mejorar las herramientas empleadas

Figura 4. Journey map analista de riesgo

Por último encontramos a Jorge, el hace parte del grupo de inversores de Sempli capacidad de prestar sus recursos a pequeñas empresas que lo requieran.

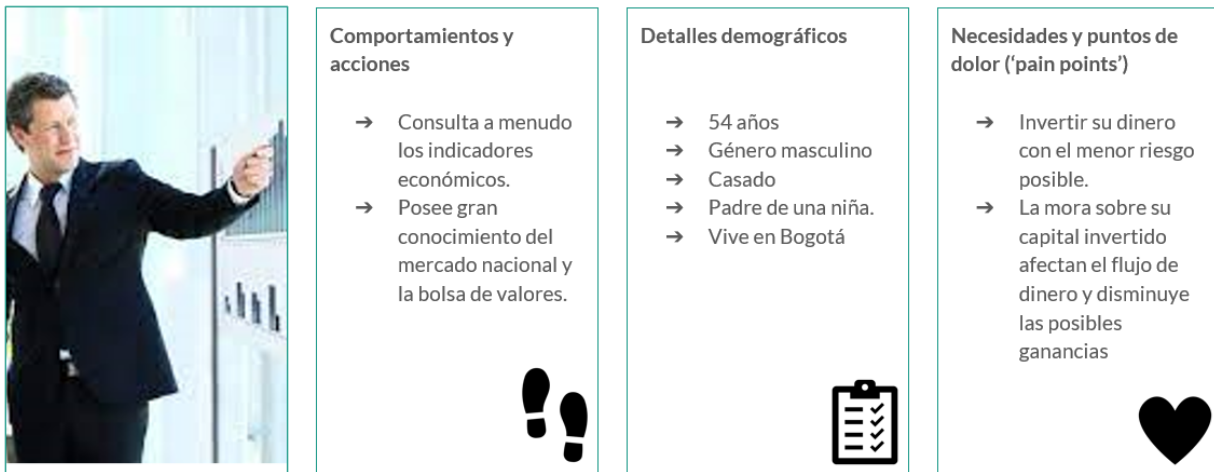


Figura 5. Arquetipo inversor

Jorge siempre está en busca de oportunidades de negocio para invertir su capital y obtener la mayor rentabilidad. El siguiente es su *journey map*.

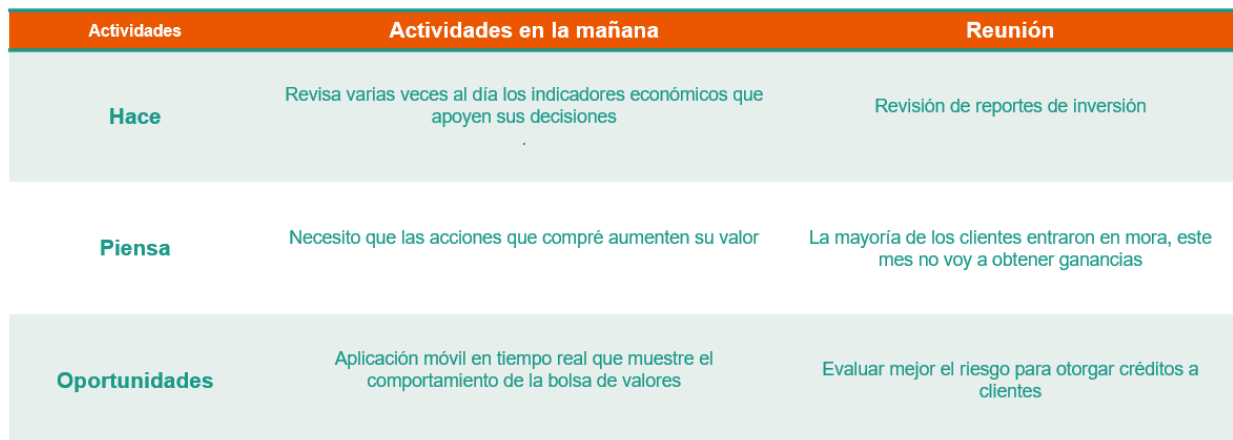


Figura 6. Journey map inversor

Hemos notado que los análisis de riesgo sobre las solicitudes de crédito tardan en dar respuesta a los clientes, esto hace que los clientes pierdan oportunidades de negocio y que Sempli pierda posibles buenos clientes. También, hemos notado que algunos clientes con buenos puntajes de riesgo posteriormente entran en mora con sus créditos. ¿Cómo podríamos evaluar, de forma rápida y eficiente, el riesgo de que los posibles clientes incumplan sus compromisos financieros con Sempli? Para dar respuesta a esta pregunta realizamos una lluvia de ideas de la cual se obtuvieron las siguientes propuestas:

- Una página web de autogestión para clientes que les permita determinar si Sempli les va a otorgar un crédito.
- Un servicio web que le permita al analista de riesgo simular rápidamente un nuevo crédito.
- Una aplicación móvil que le permita al inversor seleccionar empresas en las que desea invertir, a partir una la lista de posibles clientes con sus características.
- Desarrollo de un RPA (*Robotic Process Automation*) que apoye a los analistas de riesgo en sus labores repetitivas.
- Un servicio web que consuma un archivo plano y clasifique rápidamente a los clientes con el fin de identificar quienes pagaran y quienes no pagaran el crédito que están solicitando.

- Una página web que le permita a los clientes, simular un crédito y calcular el monto máximo que le puede otorgar Sempli basado en sus características.

De las anteriores ideas nos inclinamos por el desarrollo de un servicio web para uso exclusivo de los analistas de riesgo, cuya entrada sea un archivo plano, cada registro de este archivo corresponde a un cliente y sus características. Basado en estos datos y la experiencia de un modelo de clasificación, la respuesta del servicio sea una etiqueta para cada uno de los clientes del archivo que indique si pagará o no el crédito. Los datos utilizados para el desarrollo de esta idea son: las características de los clientes, como lo son el número de empleados, su sector económico, la ciudad y la edad de la empresa. Así como también se hará uso de los datos de los créditos ya otorgados por Sempli, estos datos incluyen: el monto prestado, la tasa de interés, el plazo o número de cuotas y el estado del crédito (finalizado, en curso, entre otros).

Con base en estos datos se pretende predecir si una empresa va a ser un buen cliente o no, soportado por un árbol de decisión que clasifique con rapidez y precisión a los clientes que soliciten un nuevo crédito a Sempli. Apoyando a los analistas de riesgo en la decisión de negar el crédito al cliente o de aprobarlo. El siguiente *story board* ilustra el funcionamiento del servicio web a desarrollar evidenciando los beneficios que obtendrán las partes interesadas:

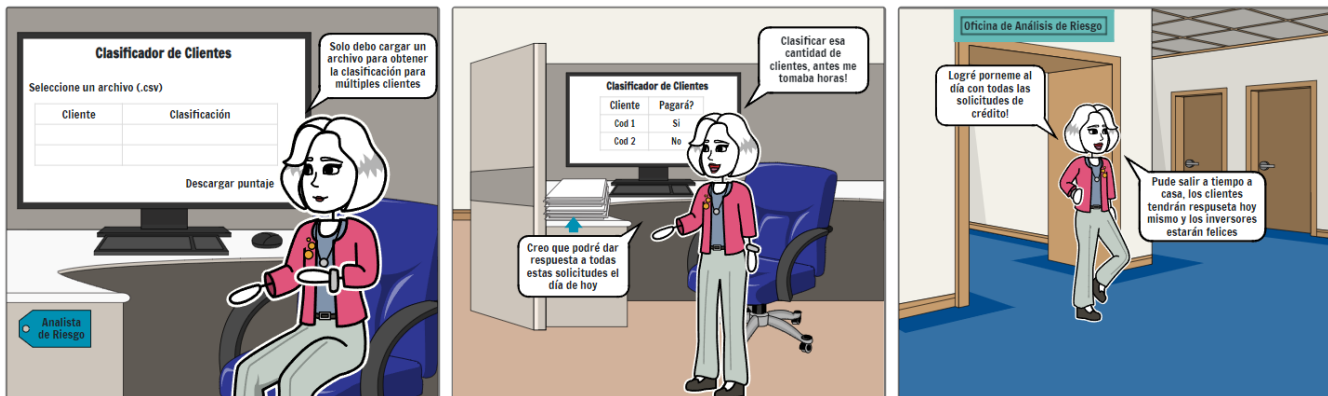


Figura 7. Story board de nuestra solución

### 3 Implicaciones Éticas

Debido a que los datos suministrados por Sempli corresponde a información real de sus clientes y productos, el equipo de trabajo tiene el compromiso de no dar acceso público a dichos datos, para no incurrir en riesgos legales y respetar la privacidad de los clientes. Para esto, antes de compartir los datos, Sempli eliminó algunas columnas del archivo con el fin de anonimizar la información.

El compromiso del equipo para el desarrollo de este proyecto, es realizar un ejercicio meramente académico que pretende predecir un puntaje de riesgo basado en características generales de las empresas. No se cuenta y no se utiliza datos específicos de las empresas como su nombre, número de identificación o información personal del gerente o representante., por lo tanto, no se desarrollarán resultados individualizados.

## 4 Enfoque Analítico

La pregunta a trabajar será: Qué perfil tienen los clientes que no pagarán de vuelta un crédito a Sempli luego de desembolsar los recursos. Esto se reduce a un modelo de clasificación que permita identificar de forma más sencilla a los clientes que no pagaran un crédito. Nuestro modelo de *machine learning* será un modelo supervisado clasificación a partir de árboles de decisión.

## 5 Entendimiento de los Datos

La compañía Sempli compartió al equipo un archivo en formato Excel con el registro de los créditos otorgados a sus clientes. Cada registro de este Excel contiene datos de la empresa (cliente) y sus créditos; información que se considera suficiente para inicialmente realizar un análisis exploratorio de datos enfocado en nuestro objetivo general de clasificar a los futuros clientes de Sempli y conocer cuáles van a cumplir sus compromisos financieros basado en sus características.

De los datos proporcionados se filtraron los créditos activos al mes de agosto del presente año (2022), se eliminaron las columnas identificador de cliente e identificador de crédito con el objetivo de seguir anonimizando la data y manualmente se ajustaron los datos de algunas columnas para una correcta lectura desde Python. Finalmente, se creó una nueva columna (cliente riesgo), con los valores dicotómicos cero (0): bajo riesgos y uno (1): alto riesgo; alto riesgo corresponde a los créditos que han tenidos dos moras o más, mayores a quince (15) días en los últimos seis (6) meses. El análisis exploratorio de datos realizado se centró en esta nueva característica.

## 6 Análisis Exploratorio de los Datos

En el Jupyter notebook adjunto, se encuentra debidamente documentado el análisis exploratorio de datos junto con los hallazgos a partir de las preguntas a trabajar previamente planteadas.

## 7 Conclusiones y Acciones

- Definimos una variable objetivo, la cual fue *cliente riesgoso*, esta variable define a un cliente según su comportamiento de pago:

1 = Clientes de alto riesgo, es decir que tienen 2 moras o más, cada una mayor a 15 días en los últimos 6 meses.

0 = Clientes de bajo riesgo, es decir que tienen menos de 2 moras, cada una mayor a 15 días en los últimos 6 meses.

- Luego de ejecutar un análisis exploratorio de datos, no encontramos una cantidad adecuada de variables con suficiente sustento estadístico para clasificar a un cliente de acuerdo a su comportamiento de pago (si pagará o no su crédito desembolsado con Sempli).

- Para continuar con el objetivo de entregar a Sempli un modelo de clasificación que permita identificar de forma más sencilla a los clientes que no pagaran un crédito, debemos continuar con la ejecución de una regresión lineal que permita encontrar las *features* que determinan la variable objetivo.