

## El problema

La empresa **La Contadora SA** se dedica a generar softwares de contabilidad (ERP o *Enterprise Resource Planning*) para pequeñas y medianas empresas. Desde el año pasado se busca aumentar el número de clientes para su producto **Contadorito**, un ERP *cloud*. Los usuarios de **Contadorito** han aumentado, pero el software ha presentado problemas y demoras en su ejecución pues no está optimizado para la cantidad de cuentas nuevas que ha tenido. El equipo de *Servicio al cliente*, a pesar de nuevas contrataciones, se ha visto sobrepasado por la cantidad de requerimientos de ayuda por parte de los usuarios, demorando mucho tiempo para dar respuesta satisfactoria e incluso dejando de contestar algunas consultas. Si bien se sabe que la solución de fondo implica un *refactor* del código, se proyecta no menos de 1 año para completarlo, por lo que es necesaria una pronta respuesta a los requerimientos que realizan los clientes y así evitar la fuga de estos.

## Metodología CRISP-DM

Se usará la metodología **CRISP-DM** para desarrollar la solución.

- **Comprensión del negocio:** Los problemas ocurren cuando aumenta el número de transacciones en el horario de mañana y con aquellos documentos que tienen muchas líneas de detalle. Dado el incremento en las consultas a *Servicio al cliente* que esta situación lleva se busca el desarrollo y despliegue de un asistente virtual en la aplicación. En una primera etapa los objetivos de este proyecto serían:

1. Disminuir los tiempos de resolución de problemas.
2. Que no exista algún requerimiento de usuarios que no sea contestado.

Para el desarrollo del proyecto se contrató 1 científico de datos y 1 ingeniero de Machine Learning quienes trabajarán junto al resto de los equipos. Se definen también los tiempos de entrega y los entregables.

- **Comprensión de los datos:** Además de los datos de la empresa (cantidad de colaboradores, giro, dirección, etc) y de los colaboradores (edad, sexo, cargo, etc), se dispone de un histórico de 2 años de transcripciones de conversaciones entre los usuarios con la gente de *Servicio al cliente*. Además, se cuenta con información técnica de la aplicación, su tráfico y sus problemas, junto con la hora en la que ocurrieron y de las soluciones que ha tenido que implementar el equipo de *Desarrollo* cuando se ha requerido.
- **Preparación de los datos:** El objetivo es tener un dataset de casos lo más completo posible con la información de los requirentes, el tipo de problema y, de ser posible, la solución: se extraerá de las conversaciones aquellos conceptos clave que referencian a módulos, acciones o documentos del ERP que presentan conflictos, de forma de categorizar los problemas que se repiten. Las áreas de *Servicio al cliente* y *Desarrollo* aportarán a cada chat el tipo de respuesta que permitió solucionar el problema.

- **Modelación:** Luego de identificar los problemas más importantes se intentará generar una clusterización que permita identificar grupos de clientes y problemas y, de ser posible, determinar si existen soluciones en común. Con esta información se podrá entrenar un modelo grande de lenguaje con pares de problemas y posibles soluciones probando distintos *prompts* para lograr respuestas certeras. Todas las conversaciones deberán poder generar un JSON con información necesaria para seguir entrenando al modelo y que pueda ser de utilidad para que una persona tome el caso si el chatbox no puede resolverla.
- **Evaluación:** Las primeras evaluaciones del modelo se harán con casos ficticios contruidos de acuerdo a ejemplos sacados de la información generada durante la clusterización. Se revisarán las conversaciones buscando respuestas adecuadas y tratando de identificar sesgos. Posteriormente se evaluarán con conversaciones generadas por personas de *Servicio al cliente* y *Ventas* quienes, además de conocer bien la aplicación para poder venderla, han recibido comentarios de los problemas de la aplicación por parte de potenciales clientes. En este punto, en caso de ser necesario, se revisarán los requerimientos iniciales y se solicitarán nuevos datos para hacer reentrenamiento de ambos modelos.
- **Explotación:** Luego de las pruebas con usuarios internos, el ingeniero de Machine Learning, junto a personal del equipo de *Desarrollo*, procederán a integrar el modelo como un chatbox en la aplicación, dejándolo disponible primero sólo para un grupo de clientes seleccionados, quienes recibirán algún beneficio por ello. Cada semana se escogerá aleatoriamente usuarios para evaluar su experiencia con el chatbot. Si se logran los objetivos anteriormente indicados en esta muestra de usuarios se liberará el bot de forma general.

Ensayo 1  
Néstor Patricio Rojas Ríos  
Curso Ciencia de datos y sus aplicaciones