

Desarrollo del Proyecto a la entidad financiera “Prymera” a ejecutarse en un tiempo tentativo de 2 meses y medio.(50 días útiles)

Plan de gestión de tiempo: 5 Fases

1. Entendimiento del Negocio:

1.1 Identificación, entendimiento y solución del problema de negocio.(3días)

-Definición de la población objetivo que tengan comportamiento en el sistema financiero no solo con la entidad sino con otras entidades.

-Filtrado de Clientes.

-Definición del Target.

¿Cómo saber si mis clientes me pagarán en los próximos meses?

Obs:

Gestionar reuniones con el cliente.

Entregable:

✓ Documentación del plan objetivo.

1.2 Recopilación de los datos y fuentes de información.(3días)

Encontrar diversas fuentes de datos relevantes y completos. ¿Cuántos datos se necesitan? Cuanto más, es mejor.

1.2.1 Historial crediticio (Rcc)

Ejemplo:

-Credit_person_dep: Es el tipo de crédito para personas dependientes que buscan atender el pago de bienes, servicios o gastos relacionados con la actividad empresarial. Este tipo de variable es categórica y son 11:

- Microconsumo
- Personal Directo
- Duplica tu Sueldo
- Convenio Descuento por Planilla
- Líneas CTS y Plazo Fijo
- Crédito Prendario
- Administrativo
- Crédito Caja Construye
- Crédito MiVivienda
- Crédito Hipotecario Vivienda
- Crédito Techo Propio

1.2.2 Información socio-económica (Reniec)

1.2.3 Información fiscal (Sunat)

1.2.4 Información censal (Censo)

1.2.5 Información vehicular

1.2.6 Información digital y de campañas (de aplicativo web y mobil)

Obs:

- Gestionar reuniones con el cliente.
- Relevamiento de información.

1.3 Selección de la muestra.(3días)

- La muestra debe estar conformada por agencias y oficinas clasificadas distribuidas en donde hay mayor incidencia de morosidad y abandono y con clientes de “Prymera” de los distintos segmentos como Micro o Pequeña empresa con comportamiento en el sistema financiero.
- ¿Con que cantidad de clientes que se conoce la morosidad y abandono en base a su data histórica nos va proporcionar? (train), y de estos ¿que cantidad son morosos y no morosos? Y ,¿qué cantidad abandonaron y no abandonaron? (Proporción de las clases)
- ¿Con que cantidad de clientes que no se conocen la morosidad y abandono nos va proporcionar para pronosticar la morosidad y abandono? (test).
- Para los clientes nuevos de “Prymera” no se tendrá mucha información, pero si de estos mismos clientes se tiene información sobre su comportamiento en otras entidades financieras, entonces se incluirá en la muestra.
- ¿Cuántos años o cuántos meses se tiene información de registros de los clientes hacia atrás?.
- Se realizará a través de la carga de dos archivos de Excel: el primero contiene información de los clientes que ya desertaron y cayeron en morosidad de la entidad bancaria (train) y el segundo el grupo de clientes que se requiere analizar. Los archivos contienen información del comportamiento de los clientes de la entidad bancaria que se utilizará para analizar el modelo.(test)
- Se seleccionan también distintos clientes para cada conjunto ,evitando que el modelo pueda reconocer en el test las características de un cliente que ya ha visto en el train ,lo cual falsearía la estimación del poder predictivo.

Entregable:

- ✓ Muestra objetivo seleccionada en archivo Excel.

1.4 Descripción, creación de variables, visualización.(10 días)

- Descripción de variables de acuerdo a su diccionario de datos.
- Preprocesamiento de los datos, etiquetado, formateo, transformación de variables cualitativas, recodificación.
- Creación de variables de acuerdo al negocio.
- Análisis exploratorio descriptivo.

Obs:

Si el cliente no cuenta con el diccionario de datos, gestionar reuniones con el cliente para la descripción de la misma.

Entregable:

- Creación de variables con su formulación en archivo Excel.

2 Preparación de los datos:

2.1 Feature Engineering (10días)

Limpieza, imputación de missing, valores atípicos, binning, transformación, one-hot, agrupaciones, escalado, correlación entre la variables.
Eliminación de la colinealidad de las variables, la distribución de las variables, en caso estas tengan distribución gaussiana la predicción será más confiable.
Se reescala la data de entrada utilizando estandarización o normalización.
Balanceo de los datos respecto a los targets para su mejor entrenamiento del modelo.
Selección de variables importantes.

Entregable:

- Variables importantes en archivo Excel.
- Dataset limpio y procesado en archivo Excel.

3. Modelado:

3.1 Construcción de los modelos (7días)

Abarca el entrenamiento del modelo. Este proceso implica "alimentar" el algoritmo con datos de entrenamiento, procesará datos y generará un modelo que puede encontrar el valor objetivo (atributo) en datos nuevos, una respuesta que desea obtener con el análisis predictivo. Se entrenarán modelos de aprendizaje supervisado y no supervisado. La elección se debe a pronosticar atributos específicos (morosidad y deserción) y agrupar objetos de datos por similitudes. Se eliminará las características recursivas (*backward selection*)

Entregable:

- Modelos en archivos .rds
- Imágenes de modelos en archivos .jpg

4. Evaluación:

4.1 Evaluación, ajustes y prueba de los modelos (3días)

Desarrollar el modelo más simple capaz de formular un valor objetivo rápido y suficientemente bien mediante ajustes del modelo. Esa es la optimización de los parámetros del modelo para lograr el mejor rendimiento de un algoritmo (validación cruzada). Luego se mostrará los resultados de las pruebas realizadas, tanto en la fase de entrenamiento como en la fase de validación. Se mostrará los porcentajes de acierto y error de cada fase y su respectiva matriz de confusión para los modelos de clasificación (Evaluación de la efectividad y precisión del modelo). Supervisión y mejoramiento de la eficacia del modelo.

Entregable:

- Imágenes de evaluación de modelos en archivos .jpg

5. Implementación

5.1 Resultados e implementación y despliegue (3días)

Esta sección es de implementación del modelo que cubre la puesta en producción de los modelos y que se comunique con otras partes del sistema y que los resultados del modelo se usen en el sistema.

Validar la correcta puesta en producción del modelo definido.

Entregable:

-Se exportara los resultados testeados de los clientes nuevos en un archivo Excel.

5.2 Seguimiento

Realizar un seguimiento al modelo que permita monitorear su performance y su adecuada integración en la gestión.

✓	1. Identificación, entendimiento y solución del problema de negocio	3dias	Semana 1
✓	2. Recopilación de los datos	5dias	Semana 1 Semana 2
✓	3. Selección de la muestra	5dias	Semana 3
✓	4. Descripción, creación de variables, visualización	10dias	Semana 3 Semana 5
✓	5. Feature Engineering	10dias	Semana 5 Semana 7
✓	6. Construcción de los modelos	7dias	Semana 7 Semana 8
✓	7. Evaluación, ajustes y prueba de los modelos	5dias	Semana 9
✓	8. Resultados e implementación y despliegue	5dias	Semana 10
	Total	50 dias	

Gestión de Recursos Humanos

	Rol	Miembro	Estrategia
1		Cliente	<ul style="list-style-type: none">-Evaluar y hacer seguimiento a los avances del proyecto en la primera fase.- Responder consultas con respecto al proyecto en la primera fase.-Disponibilidad de reuniones con respecto a los datos proporcionados.- Aprueba los resultados del proyecto.
2		Yesin Camarena	<ul style="list-style-type: none">- Definición del alcance.- Gestión del plan del proyecto.- Seguimiento del cronograma.- Reuniones con el cliente.- Apoyo en la investigación y validación de la documentación.- Coordinación con las autoridades involucradas en el proyecto.-Cumplir con los entregables.-Desarrollo.

Gestión de Riesgos:

En esta sección se menciona los riesgos que pueden suceder en el transcurso del proyecto. La probabilidad se medirá del 0% al 100%. El impacto se medirá con tres valores: Alto, medio y bajo.

	Riesgo	Probabilidad	Impacto	Estrategia
1	Escasa disponibilidad del Cliente para las reuniones pactadas.	10%	Alto	Pactar anticipadamente las fechas de las reuniones y/o en caso ocurra un imprevisto reprogramar instantáneamente la reunión.
2	Constante cambio del alcance del proyecto.	20%	Alto	Mantener comunicación con el cliente y la gerencia del desarrollo del proyecto para que esté alineado al alcance y objetivos.
3	No contar con grandes cantidades de información sobre las variables a considerar en el modelo.	20%	Medio	Se trabajará con la información de la entidad financiera en específico sin identificación del cliente, pero sí sobre su comportamiento financiero, demográfico, personal, etc.