

UNIVERSIDAD DEL VALLE DE GUATEMALA

Facultad de Ingeniería

CC3100 – Machine Learning Operations

Sección 20

Ing. Kevin Vidal



MLOps con CRISP-DM.

Proyecto Final

Diego Alonzo- 20172

Eduardo Ramírez - 19946

GUATEMALA, 10 de octubre de 2025

Comprensión del Negocio (Business Understanding)

Problemática de negocio (General)

Predicción de abandono de clientes de llamadas en espera en el sector de cobros dentro de un Banco debido a la alta demanda clientes en llamada en horario de oficina. Asimismo, el banco afronta una baja recuperación de dinero al no poder alcanzar a todos los clientes que están en mora.

Descripción de la empresa

Nombre de la empresa	Banco Estrellitas
Sector	Financiero
Ubicación	Guatemala, zona 6 de la ciudad capital
Exposición total	1×10^{30}

Contexto

El Banco Estrellitas afronta una crisis en su sector de cobranza, actualmente se logra recuperar únicamente el 20% de sus clientes que entran en mora, significando que el 80% restante termina siendo irrecuperable.

La principal razón es debido a que en su Call Center de parte del sector de cobranza se pierde el contacto con cientos de clientes que representan la mayor parte de su exposición (exposición es la cantidad de dinero que el banco almacena para salvaguardar una deuda por ley). La plataforma *Nube Comienzo*, es un manejador de llamadas, sin embargo, cuándo todas las líneas están ocupadas mantiene a sus clientes en espera. A pesar de esto se pierde el contacto con los clientes debido a que abandonan la llamada, implicando la pérdida de la oportunidad de lograr realizar algún acuerdo o negociar con los clientes para poder recuperar cierta parte del dinero en mora.

Por lo tanto, la dirección de la empresa ha decidido invertir en un proyecto de Machine Learning Operations (MLOps) para automatizar el monitoreo y la predicción de la probabilidad de abandono de la llamada. Permitiendo la retención del cliente y por ende aumentando el volumen de clientes con los cuáles se puede negociar y recuperar el dinero que está en mora y expuesto. Se denota el hecho de que no se poseen estrategias de retención de clientes y que disponen de un precio de 2 millones de dólares (alrededor de 15 millones de quetzales).

Problema a resolver

La empresa no posee un sistema predictivo para identificar qué clientes tienen la probabilidad más alta de abandonar la llamada. Actualmente, no se disponen de estrategias de retención de clientes por lo que este sería un primer paso para la misma.

Impacto económico del problema

La continua pérdida del 80% de sus clientes en mora, de forma que el dinero que está expuesto se convierte en dinero irrecuperable. Lo que podría derivar en que el banco cierre ya que al mantener ese dinero guardado no lo puede gestionar para realizar más préstamos y poder nutrir sus finanzas. Se aproxima una pérdida de Q. 1×10^{30} , por lo que es un problema grave, dado que podría significar el cierre del banco en caso de que no sea solvente.

El costo de oportunidad en este caso serían 15 millones de quetzales que no podrían utilizarse en préstamos o en manejo de dinero para generar más dinero.

Stakeholders involucrados

Dentro de este proyecto existen varios stakeholders a diferentes niveles:

- Stakeholder a nivel de call center, de forma que un jefe dentro del call center de cobros pueda saber cuándo un cliente disponga de una probabilidad mayor de abandono de forma que se le de cierta prioridad. Por lo que necesita conocer aquellos clientes que tengan una probabilidad mayor de abandono para disminuir el mismo.
- Stakeholder a nivel de dirección general, un CEO necesita poder aumentar la recuperación de dinero del sector de cobranza. De forma que pueda tomar decisiones estratégicas.
- Stakeholder a nivel de gerencia, poder identificar aquellos clientes que estén en riesgo para focalizarlos y poder establecer una estrategia específica para ellos.
- Stakeholder a nivel de equipo de ciencia de datos, poder contar con un pipeline de MLOps reproducible, automatizado y que permite monitorizar el modelo.
- Stakeholder a nivel de TI, permite disponibilizar el modelo a partir de infraestructura que soporte la disponibilidad instantánea del modelo.

Restricciones técnicas

1. Infraestructura On-Premise

No se permite el uso de servicios de la nube por políticas de seguridad bancaria, asimismo, requiere de un hardware local robusto así como servidores con GPU

para modelos. Además, existe una mayor complejidad en la escalabilidad y mantenimiento, ya que se debe de configurar redes internas seguras para el despliegue de API's y servicios.

2. MLFlow para MLOps

Se requiere que MLFlow sea gestionado e instalado de forma local. Asimismo, se debe de configurar un servidor MLFlow para la infraestructura del banco. Requiere almacenamiento local para modelos, métricas y artefactos.

3. Tiempo

Se contempla el desarrollo del modelo en 12 semanas dada la complejidad de los datos y la masividad de los datos a utilizar. Por lo que tanto las fases de entrenamiento, de prueba, limpieza, e ingeniería tomarán más tiempo al requerir de un análisis inicial de datos y exploración de los mismos.

4. Presupuesto

Se dispone de un presupuesto de 15 millones de quetzales para poder montar todo el proyecto.

Screenshots

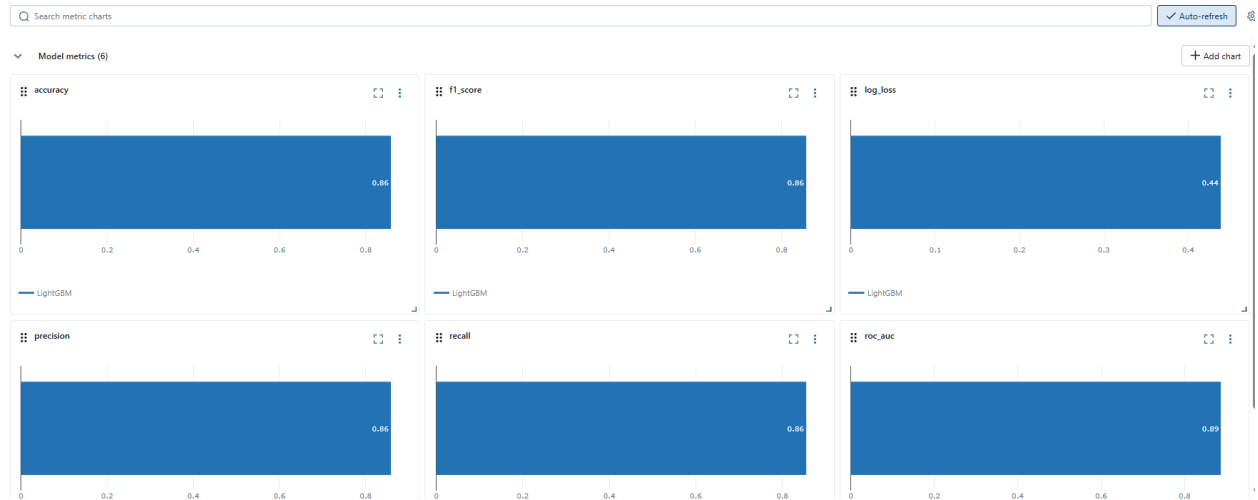
Experimentación con múltiples modelos

<input type="checkbox"/>	Run Name	Created	Duration	accuracy	f1_score	log_loss	precision	recall	roc_auc	learning_rate	max_depth	n_estimators	n_features	n_samples_train	n_samples_val
<input type="checkbox"/>	LightGBM	29 minutes ago	7.6s	0.85844748...	0.85699222...	0.43871333...	0.85866991...	0.85530123...	0.886683463...	0.2	3	50	2239	14863	7665
<input type="checkbox"/>	XGBoost	30 minutes ago	8.8s	0.86027397...	0.85898617...	0.43672230...	0.85977859...	0.85819521...	0.88814646...	0.1	3	100	2239	14863	7665
<input type="checkbox"/>	RandomForest	32 minutes ago	8.5s	0.84918460...	0.84821428...	0.41824514...	0.84665792...	0.84977637...	0.88254446...	-	20	200	2239	14863	7665
<input type="checkbox"/>	LogisticRegression	47 minutes ago	17.4s	0.82191780...	0.82173174...	0.46031930...	0.81587136...	0.82767692...	0.87573057...	-	-	-	2239	14863	7665
<input type="checkbox"/>	LogisticRegression	1 hour ago	2.2s	-	-	-	-	-	-	-	-	-	-	-	-
<input type="checkbox"/>	LightGBM	1 hour ago	5.3s	0.85844748...	0.85699222...	0.43871333...	0.85866991...	0.85530123...	0.886683463...	0.2	3	50	2239	14863	7665
<input type="checkbox"/>	XGBoost	1 hour ago	12.1s	0.86027397...	0.85898617...	0.43672230...	0.85977859...	0.85819521...	0.88814646...	0.1	3	100	2239	14863	7665
<input type="checkbox"/>	RandomForest	1 hour ago	5.7s	0.84918460...	0.84821428...	0.41824514...	0.84665792...	0.84977637...	0.88254459...	-	20	200	2239	14863	7665
<input type="checkbox"/>	LogisticRegression	1 hour ago	13.8s	0.82191780...	0.82173174...	0.46031930...	0.81587136...	0.82767692...	0.87573057...	-	-	-	2239	14863	7665

Screenshots integrados en la presentación

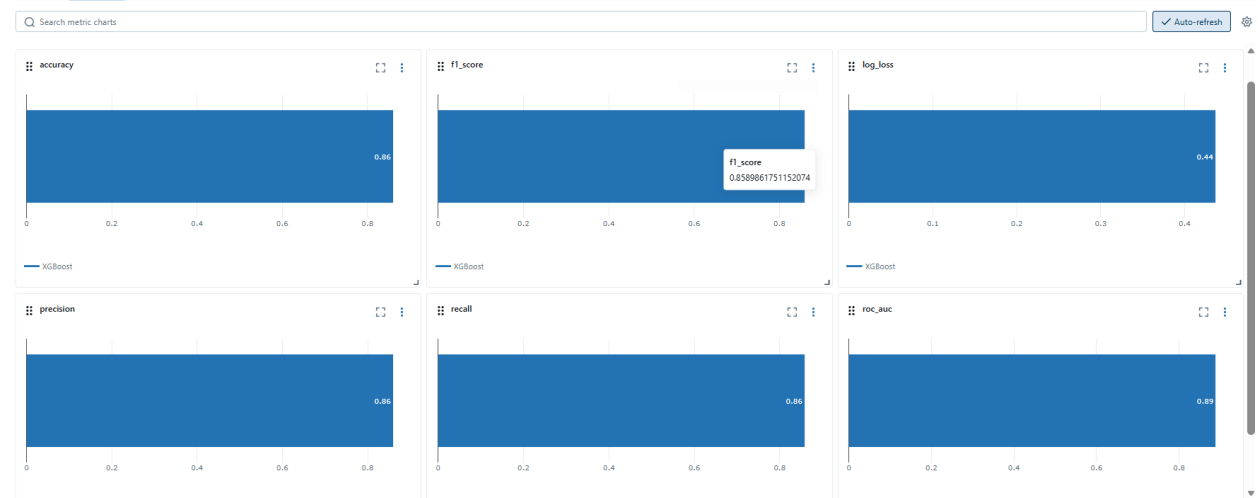
LightGBM

Overview **Model metrics** System metrics Traces Artifacts



XGBoost

Overview **Model metrics** System metrics Traces Artifacts



RandomForest

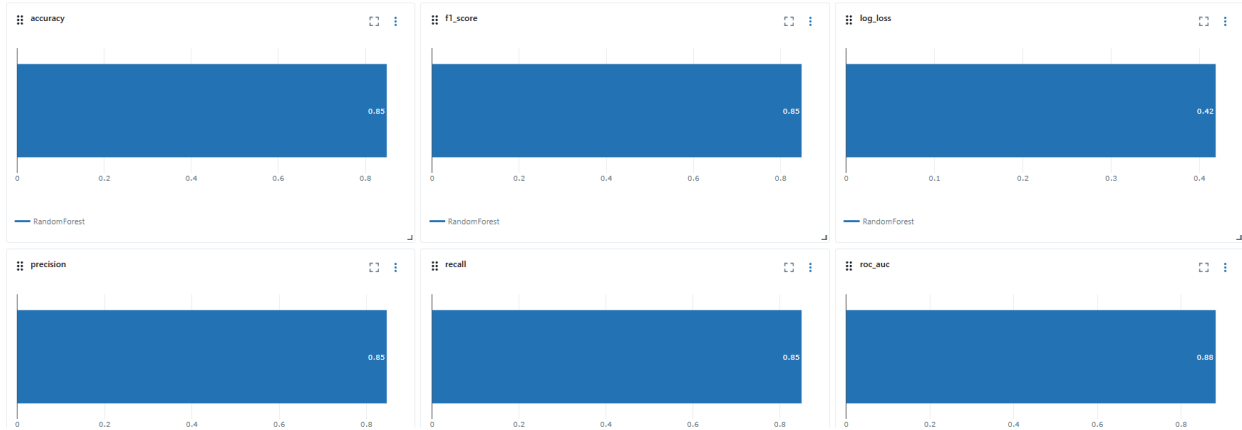
Overview **Model metrics** System metrics Traces Artifacts

Q Search metric charts

✓ Auto-refresh

Model metrics (6)

+ Add chart



LogisticRegression

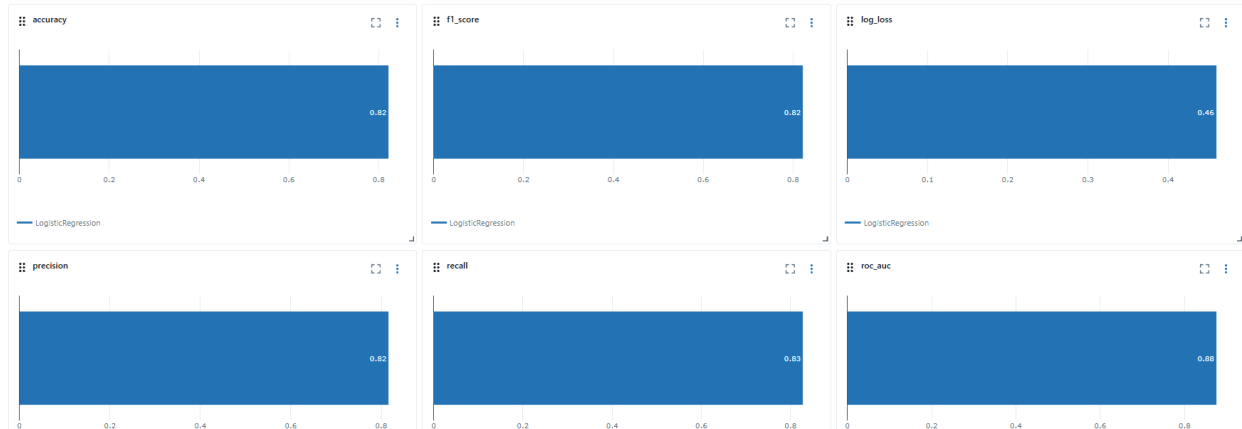
Overview **Model metrics** System metrics Traces Artifacts

Q Search metric charts

✓ Auto-refresh

Model metrics (6)

+ Add chart




churn-prediction-experiment > Runs >

LogisticRegression

Overview Model metrics System metrics Traces **Artifacts**

 cv_results_logistic_regression.json

 hyperparameter_space_logistic_regression.json

hyperparameter_space_logistic_regression.json 216B


Path: file:///C:/Users/diego/MLOPS-Final-Proyect/mlruns/1/a88960b296fd4454b4619f0b5ef5a764/artifact


```
{
  "C": [
    0.001,
    0.01,
    0.1,
    1,
    10,
    100
  ],
  "penalty": [
    "l1",
    "l2"
  ],
  "solver": [
    "liblinear",
    "saga"
  ],
  "max_iter": [
    1000,
    2000
  ]
}
```


churn-prediction-experiment > Runs >

LightGBM

Overview Model metrics System metrics Traces **Artifacts**

 cv_results_lightgbm.json

 hyperparameter_space_lightgbm.json

hyperparameter_space_lightgbm.json 478B


Path: file:///C:/Users/diego/MLOPS-Final-Proyect/mlruns/1/e10dba6470f64810a16ea


```
{
  "n_estimators": [
    50,
    100,
    200,
    300
  ],
  "max_depth": [
    3,
    5,
    7,
    9,
    -1
  ],
  "learning_rate": [
    0.01,
    0.1,
    0.2,
    0.3
  ],
  "subsample": [
    0.6,
    0.8,
    1
  ],
  "colsample_bytree": [
    0.6,
    0.8,
    1
  ],
  "num_leaves": [
    31,
    50,
    100,
```

churn-prediction-experiment > Runs >

XGBoost

Overview Model metrics System metrics Traces **Artifacts**

 cv_results_xgboost.json

 hyperparameter_space_xgboost.json

hyperparameter_space_xgboost.json 351B

Path: file:///C:/Users/diego/MLOPS-Final-Proyect/mlruns/1/5cc87e5e

```
{
  "n_estimators": [
    50,
    100,
    200,
    300
  ],
  "max_depth": [
    3,
    5,
    7,
    9
  ],
  "learning_rate": [
    0.01,
    0.1,
    0.2,
    0.3
  ],
  "subsample": [
    0.6,
    0.8,
    1
  ],
  "colsample_bytree": [
    0.6,
    0.8,
    1
  ],
  "gamma": [
    0,
    0.1,
    0.2
  ]
}
```

churn-prediction-experiment > Runs >

RandomForest

Overview Model metrics System metrics Traces **Artifacts**

cv_results_random_forest.json

hyperparameter_space_random_forest.json

hyperparameter_space_random_forest.json 310B

Path: file:///C:/Users/diego/MLOPS-Final-Proyect/mlruns/1/6e6fcb07182c

```
{
  "n_estimators": [
    50,
    100,
    200,
    300
  ],
  "max_depth": [
    5,
    10,
    15,
    20,
    null
  ],
  "min_samples_split": [
    2,
    5,
    10
  ],
  "min_samples_leaf": [
    1,
    2,
    4
  ],
  "max_features": [
    "sqrt",
    "log2",
    null
  ]
}
```

Repositorio

<https://github.com/DiggsPapu/MLOPS-Final-Proyect.git>