**1. Propuesta de Valor**

El presente proyecto tiene como objetivo anticipar el deterioro crediticio de la cartera del banco, específicamente la evolución de los saldos en Stage 2 y Stage 3, de acuerdo con la normativa contable IFRS 9. En la actualidad, el seguimiento del deterioro se realiza de forma retrospectiva, lo que dificulta una respuesta temprana y eficiente ante señales de riesgo creciente.

Este proyecto propone el uso de modelos de series temporales basados en arquitecturas de Deep Learning (LSTM y Transformer) para predecir la proporción de deuda deteriorada (Stage 2 y 3) por combinación de Destino crediticio y Moneda. La anticipación de estos movimientos permitirá al banco:

* **Mejorar la estimación de provisiones contables**, al contar con una proyección más precisa de la evolución futura del riesgo crediticio.
* **Optimizar el uso de capital**, al evitar sobre o subestimaciones del Expected Credit Loss (ECL) y ajustarse de forma más eficaz a los requerimientos regulatorios.
* **Tomar decisiones comerciales proactivas**, como reestructuración de líneas, segmentación de campañas, adecuación de límites o refinanciaciones, focalizadas en los productos y destinos más expuestos al deterioro.
* **Fortalecer la planificación financiera y la gestión de riesgos**, ya que los outputs del modelo pueden integrarse a dashboards de monitoreo y alertas tempranas.

Además, al utilizar datos que provienen del Data Warehouse institucional del banco, con registros desde 2023 hasta Junio 2025, se garantiza una cobertura amplia de situaciones económicas (ciclos alcistas, recesivos, devaluaciones, entre otros), lo cual brinda solidez y robustez al entrenamiento del modelo.

**2. Dataset**

Se utilizará un dataset proveniente del Data Warehouse del banco, con actualización mensual y validación institucional, que contiene información agregada a nivel mensual sobre:

* **Moneda**
* **Destino del crédito**
* **Stage (1, 2 o 3)**
* **Segmento de cliente**
* **Deuda total en pesos/dólares por combinación**

El periodo de análisis abarca desde enero 2023 hasta junio 2025, lo cual permite construir series históricas completas para entrenar modelos predictivos robustos. A partir de esta base, se calcula la proporción de deuda deteriorada (Stage 2 y 3) como variable objetivo.

**Regresores utilizados**:

* **Indicadores de calendario**: mes\_sin y mes\_cos (codificación cíclica del mes para capturar estacionalidad).
* **Inflación mensual (IPC/INDEC)**: aumento de precios que tensiona la mora y el deterioro.
* **RIPTE (salarios registrados)**: proxy del ingreso real del sector privado de los hogares; cuando el salario real cae, suele aumentar el deterioro (Stage 2+3).

Se analizaron siete combinaciones de moneda y destino seleccionadas como segmentos prioritarios.

**3. Metodología de Trabajo**

El abordaje metodológico consistió en una secuencia estructurada de etapas que permitieron transformar los datos crudos en predicciones útiles para la toma de decisiones:

1. **Preprocesamiento y transformación de datos**Se agruparon los datos mensuales por Periodo, Moneda y Destino, calculando la variable objetivo como la proporción de deuda en Stage 2 y 3 sobre el total. Además, se incorporaron regresores temporales y macroeconómicos (ej. inflación mensual).
2. **Construcción de ventanas temporales (sliding windows)**Se estructuraron las series como ventanas supervisadas de entrada (*lookback* de 12 meses) y salida (*horizon* de 6 meses), permitiendo la adaptación de modelos secuenciales a tareas de forecasting.
3. **Entrenamiento y validación por segmento**Para cada combinación de Moneda y Destino, se entrenaron modelos independientes utilizando un 80% de los datos para entrenamiento y un 20% para validación. Se emplearon técnicas de *early stopping* y normalización para asegurar la estabilidad y generalización del aprendizaje.
4. **Incorporación de regresores exógenos**: se construye una matriz de covariables con mes\_sin, mes\_cos, inflación y RIPTE, alineadas al índice mensual de la serie objetivo (Stage 2+3/Total). Se evita *data leakage* ajustando los escaladores únicamente con datos de entrenamiento y aplicándolos luego a validación y horizonte futuro.
5. **Predicción a futuro (forecast)**Finalizado el entrenamiento, se generaron predicciones hacia los siguientes 6 meses, permitiendo observar la evolución esperada del deterioro crediticio sin necesidad de datos reales futuros.
6. **Evaluación y visualización de resultados**Se evaluó el rendimiento mediante las métricas MAE (Mean Absolute Error) y MSE (Mean Squared Error), seleccionadas por su capacidad de interpretar errores medios y penalizar grandes desvíos. Las predicciones fueron visualizadas junto al histórico para facilitar el análisis comparativo.

**4. Modelos a Entrenar**

Se entrenan dos modelos secuenciales:

* **LSTMForecaster**: modelo de tipo Long Short-Term Memory, especializado en capturar dependencias temporales de corto y mediano plazo en series secuenciales. Es útil para detectar dinámicas recurrentes o estacionales en los datos.
* **TransformerForecaster**: modelo basado en atención, con mayor capacidad para capturar relaciones a largo plazo entre variables temporales. Resulta más potente en contextos con múltiples factores y estructuras de tiempo no lineales.

Ambos modelos se entrenan sobre series desagregadas por **Moneda y Destino**, sin agrupar por segmento, para una mayor granularidad y control en la toma de decisiones.

**5. Métricas de Evaluación**

Las métricas utilizadas para evaluar el rendimiento de los modelos son:

* **MAE (Mean Absolute Error)**: mide el error medio absoluto entre la predicción y el valor real. Es interpretada en las mismas unidades que la variable objetivo (proporción de Stage 2 y 3). Es fácil de interpretar y robusta ante valores atípicos, por eso es útil cuando se busca entender cuánto nos desviamos, en promedio, al hacer una predicción.
* **MSE (Mean Squared Error)**: mide el error cuadrático medio, penalizando más fuertemente errores grandes. Se usa en este contexto para captar posibles errores significativos que puedan tener mayor impacto en el capital regulatorio y el balance.

Ambas métricas son adecuadas para este tipo de problemas porque permiten cuantificar con claridad el desempeño del modelo al predecir proporciones, y porque las diferencias pequeñas en Stage 2/3 pueden tener efectos relevantes en las previsiones contables del banco.