

ABSTRACT

Este estudio explora la aplicación de algoritmos de Machine Learning para acelerar el cálculo del riesgo de crédito en contratos derivados, con un enfoque particular en el Credit Value Adjustment (CVA). Los hallazgos demuestran mejoras significativas en los tiempos de ejecución manteniendo la precisión, lo que potencia una gestión de riesgo más eficiente en el sector financiero.

PROBLEMA DE NEGOCIO

Necesidad: Con el crecimiento del mercado financiero, se incrementa la necesidad de cuantificar el riesgo de crédito y las exposiciones de los contratos derivados. El cálculo del CVA es computacionalmente intensivo, por lo que, para manejar el volumen creciente de transacciones y permitir soluciones en tiempo real, es fundamental acelerar su ejecución.

Objetivo: Calcular el CVA de los FX forwards con algoritmos de Machine Learning de manera instantánea.

Credit Value Adjustment (CVA): El precio de un contrato FX forward en el tiempo $t = 0$ está dado por $V(0, T) = Q(0)e^{(r_2(0, T) - r_1(0, T))T}$ donde

- T es el tiempo de maduración,
- $Q(0)$ denota la tasa de cambio entre la divisa D_1 y la divisa D_2 ,
- $r_1(0, T)$ y $r_2(0, T)$ denotan la tasa de interés de la divisa D_1 y la divisa D_2 respectivamente (de $t = 0$ a maduración).

El CVA para un FX forward se calcula generando M simulaciones de las variables de mercado para aproximar la exposición positiva esperada:

$$EPE(t) = \frac{1}{M} \sum_{m=1}^M \text{descuento}_m(t) \times \max[V_m(t, T), 0]$$

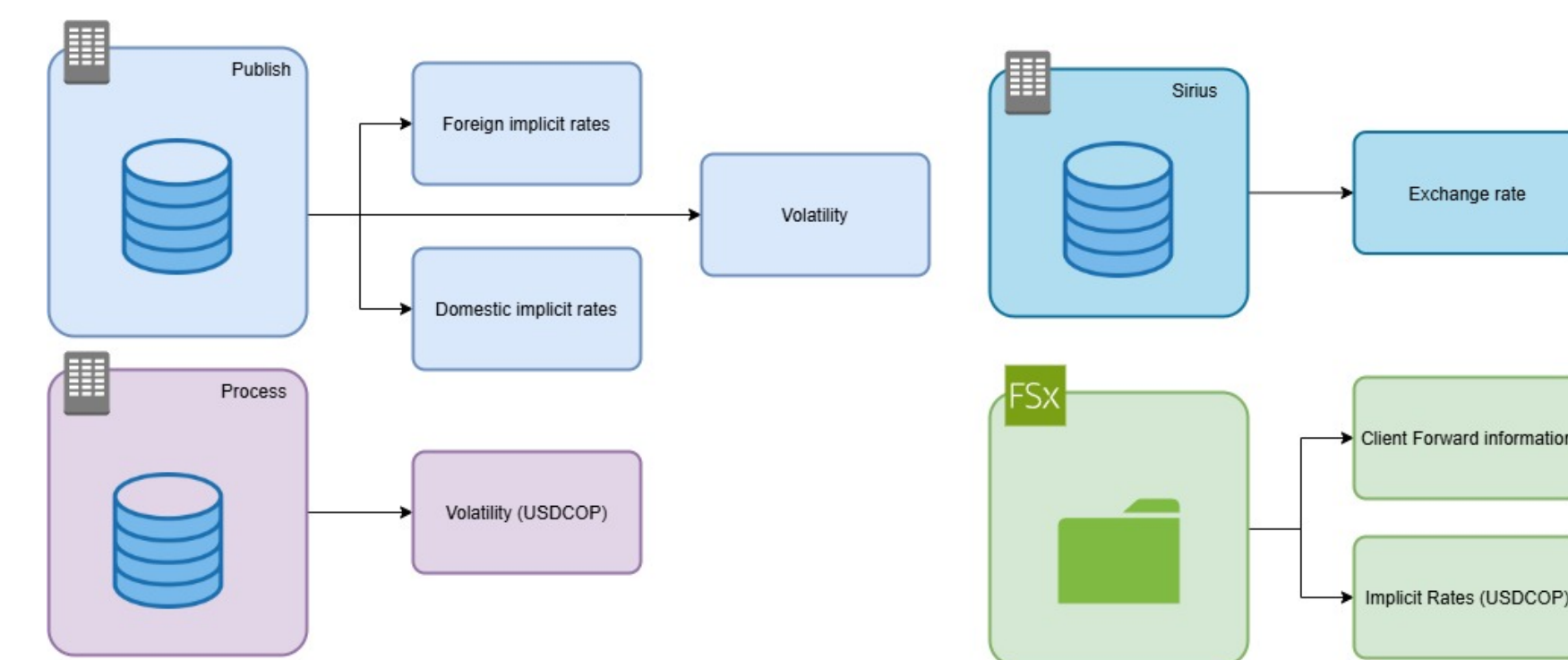
Con la exposición positiva esperada podemos calcular el dinero que se pierde si la contraparte se quiebra de la siguiente manera:

$$CVA = (1 - R) \sum_{k=1}^N EPE(t_k) PD(t_{k-1}, t_k)$$

donde R es la tasa de recuperación y $PD(t_{k-1}, t_k)$ es la probabilidad de default de la contraparte entre el tiempo t_{k-1} y t_k .

INGENIERÍA DE DATOS

El objetivo principal de esta etapa fue asegurar la integración adecuada y la preparación detallada de múltiples fuentes de información relevantes para la predicción del Credit Value Adjustment (CVA). Los datos provienen de bases como aurora-publish, aurora-process, sirius (RDS), y FSx (servidor de archivos).



Los procesos clave realizados incluyen:

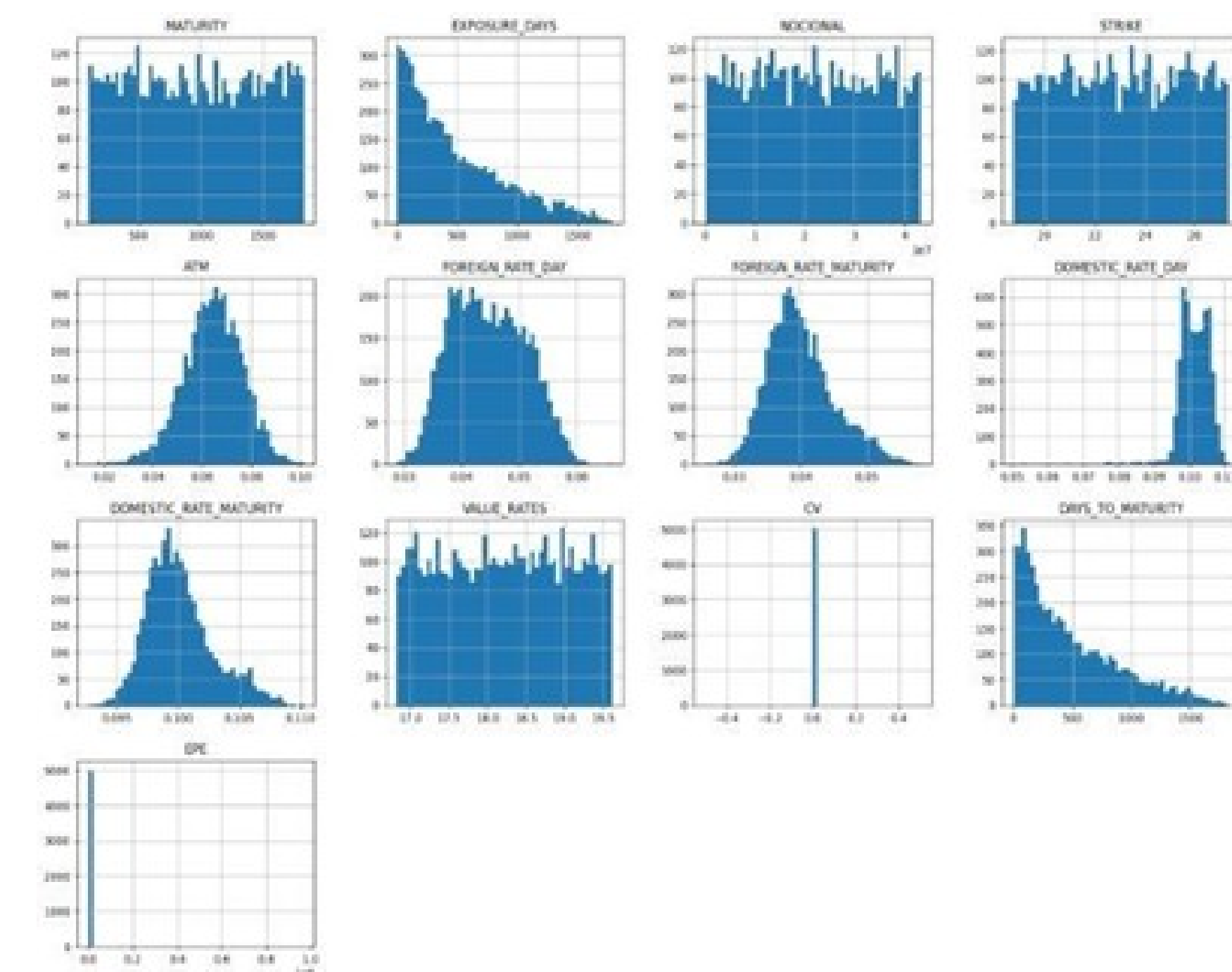
Extracción de Datos: Consultas SQL sobre RDS y lectura desde FSx.

Selección de características: Selección iterativa de variables significativas y exclusión de columnas redundantes.

Transformación y Uniformización: Preparación minuciosa de datos para asegurar coherencia y utilidad analítica en dos iteraciones claramente diferenciadas:

- *Primera Iteración (Archivos Planos):* Integración manual mediante múltiples joins y transposición de estructuras.
- *Segunda Iteración (Bases de Datos):* Optimización con consultas dinámicas y creación de variables derivadas como *RATE_DIFF*, *DISCOUNT*, *EXPONENTIAL_TERM*.

Sampleo de Datos: Generación sintética para asegurar diversidad y representatividad en el entrenamiento del modelo predictivo.



ENTRENAMIENTO Y PREDICCIÓN

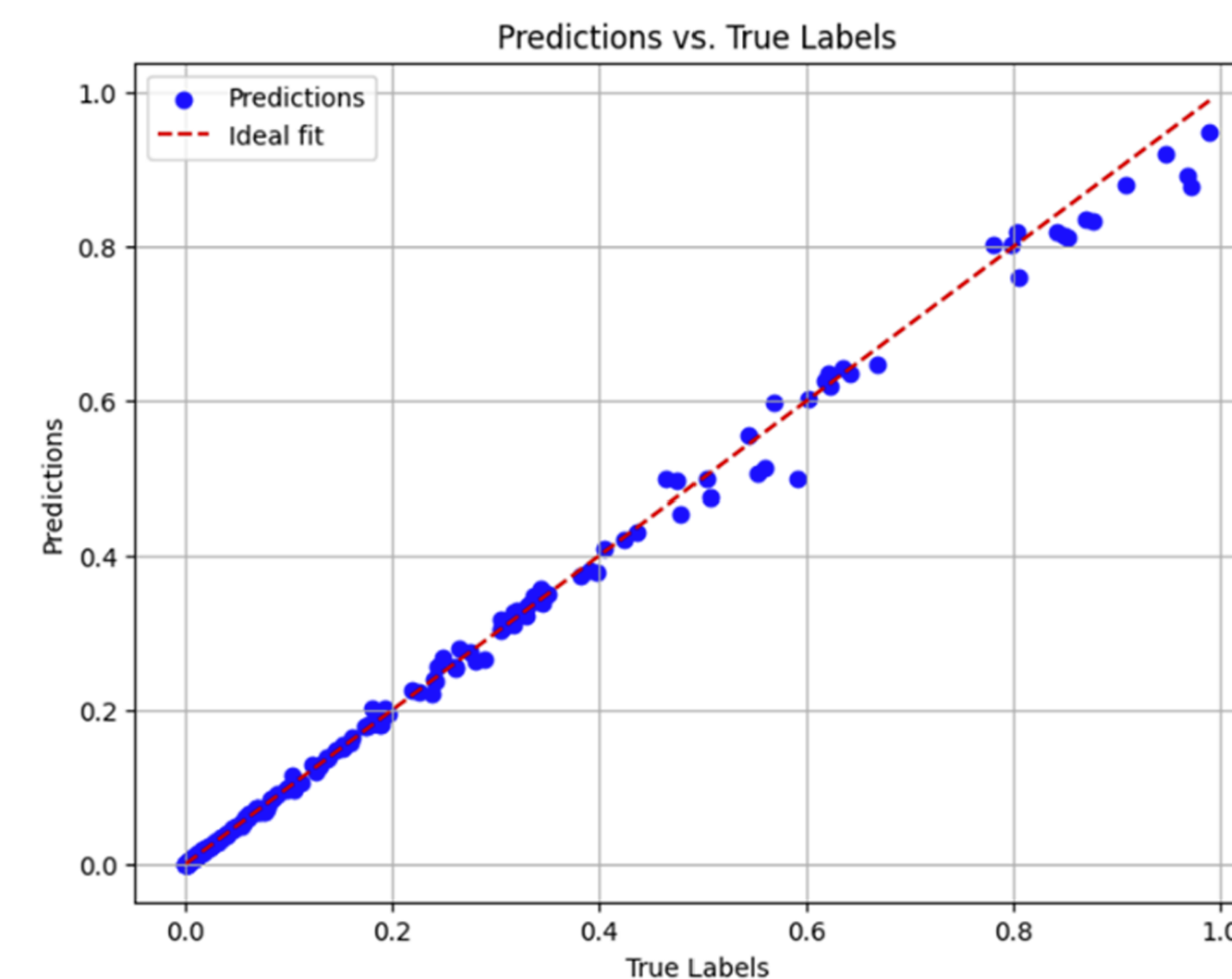
Proceso de entrenamiento y tuning : Python, scikit-learn, Pytorch, W&B y Ray.

- **Entrenamiento:** train-test split.
- **Tuning:** k-fold cross-validation.

Modelos utilizados: El modelo que obtuvo los mejores resultados fue el RandomForest.

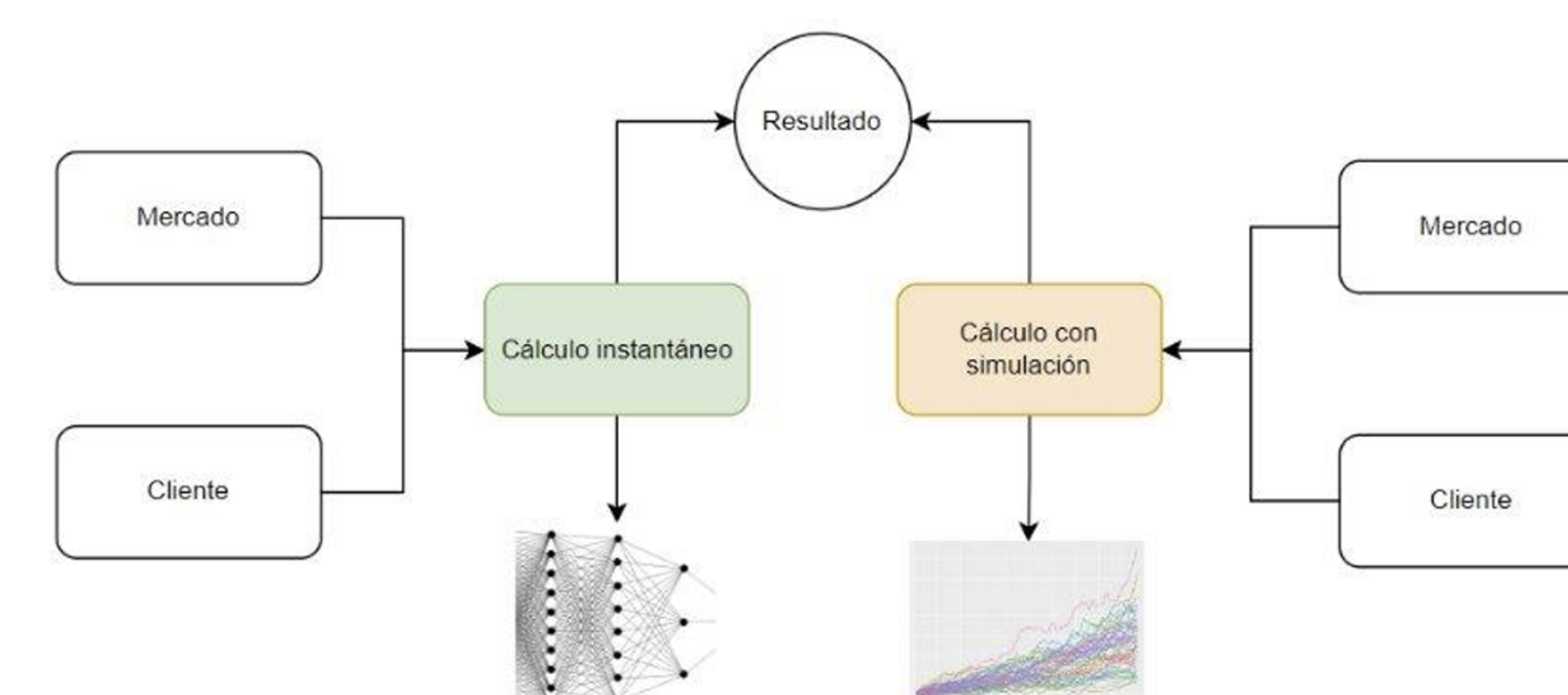
- RandomForest.
- XgBoost.
- Deep Learning.

Resultados: MAE del 0.4% en la predicción de la exposición esperada (EPE) para FX forward USD-MXN.



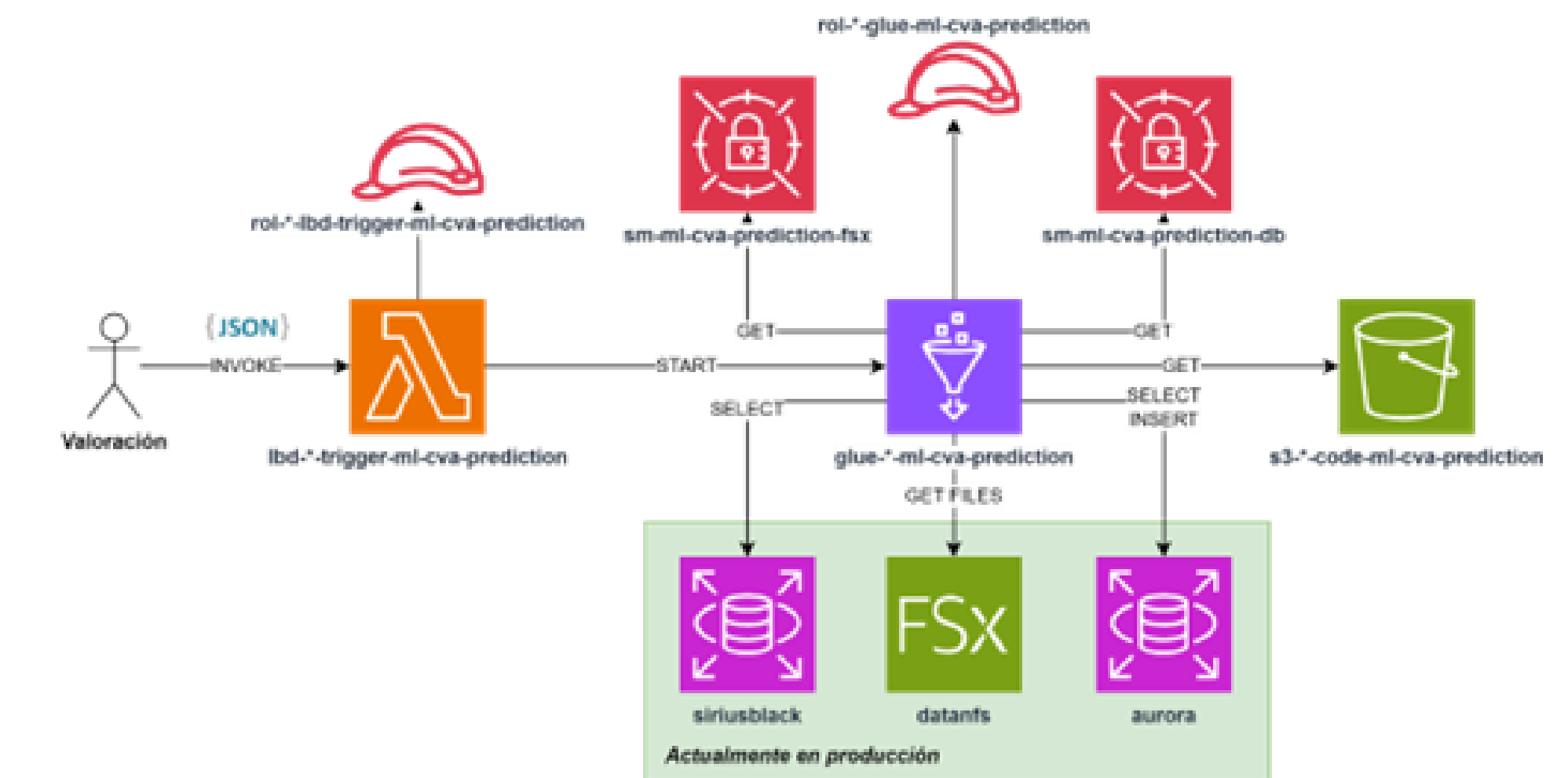
Proceso de predicción:

1. Obtener datos.
2. Predecir las exposiciones esperadas con el estimador entrenado.
3. Utilizar la fórmula del CVA con las exposiciones esperadas obtenidas en el literal 2.



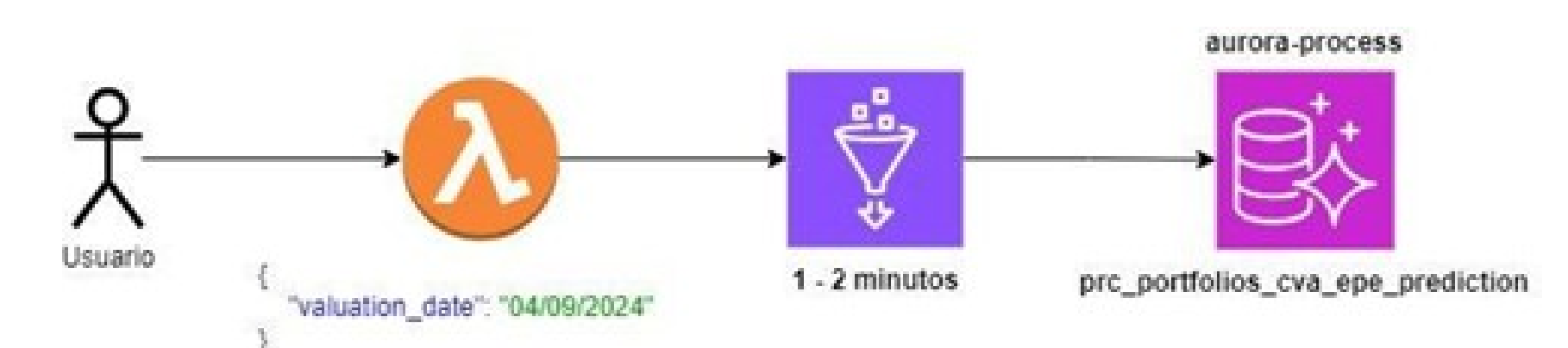
DESPLIEGUE

Una vez entrenado y evaluado el modelo, se desplegó una arquitectura en la nube utilizando servicios de Amazon AWS que ya hacen parte del ecosistema operativo de la empresa. La solución combina funciones Lambda, orquestación con AWS Glue, y almacenamiento de modelos en S3.



Flujo de ejecución:

1. El usuario activa la predicción enviando una fecha de evaluación (*valuation_date*) a través de AWS Lambda.
2. Glue extrae datos desde RDS y FSx, transforma la información, ejecuta el modelo almacenado en S3 y genera las predicciones.
3. Los resultados se almacenan en la base de datos Aurora bajo el esquema *prc_portfolios*, tabla *prc_portfolios_cva_epe_prediction*.



Reentrenamiento: El proceso contempla reentrenamiento local en notebooks compatibles con el entorno de Glue, ajustando los rangos de simulación y actualizando los modelos .pkl almacenados en S3.

CONTACTO

Nicolás Suarez, Head of R&D, Precia PPV S.A. Estudiante Doctoral, Management (Finance), Universidad de Los Andes. **Correo:** nsuarez@precia.co

Esteban Leiva, MSc Ingeniería Industrial, Universidad de Los Andes. **Correo:** e.leivam@uniandes.edu.co

Santiago Martínez, MSc Ingeniería de Información, Universidad de Los Andes. **Correo:** s.martinezn@uniandes.edu.co