

TP Deep Learning

- Análisis de Correlación de Consumos TC y Saldos Vista
- Predicción de Saldos Vista

Agosto 2025

Andrés Beneditto

Contexto y Objetivo del Proyecto

Objetivo del Análisis:

- Comprender la relación entre los patrones de consumo (tarjetas de crédito/débito) y la evolución de los saldos en cuentas vista (saldos en Caja de Ahorro de los Clientes).
- Construir un modelo predictivo para pronosticar los futuros saldos promedio en cuentas vista, utilizando todas las variables económicas disponibles.

Relevancia: Predecir los saldos permite anticipar la liquidez de los clientes, optimizar la oferta de productos financieros y gestionar el riesgo.

Variable Objetivo: El foco del pronóstico es el Saldo Promedio Vista, ajustado por inflación (IPC) para un análisis más preciso.

Dataset

El Dataset consiste en 30.000 clientes seleccionados al azar (a través de consulta SQL al DataWarehouse de la entidad Bancaria) de cada Segmento (Jubilados, Plan Sueldo Privado, Plan Sueldo Público). Un total de 90.000 clientes. Estos clientes se repiten a lo largo de todos los periodos, que abarcan desde Diciembre 2022 a Mayo 2025.

Fuentes de Datos Combinadas:

- Datos Transaccionales del Cliente: Información interna que incluye consumos, saldos y tenencia de productos.
- Datos Macroeconómicos: Se incorporó el Índice de Precios al Consumidor (IPC) de fuentes públicas para contextualizar el entorno económico.

Variables Clave Utilizadas:

- Variable Objetivo: saldo_promedio_vista_ipc (Saldo promedio del Mes en Caja de Ahorro, deflactado).
- Principales Features (Predictores):
 - Consumos con tarjeta de débito y crédito (Consumo TD, Consumo TC).
 - Nivel de ingresos del cliente (Acreditacion Salarial).
 - Índice de tenencia de otros productos (Cross Sell).

Tratamiento Esencial de los Datos:

- Ajuste por Inflación: Todas las variables monetarias fueron deflactadas utilizando el IPC. Este paso es crucial para eliminar distorsiones y permitir una comparación justa de los valores en el tiempo.
- Dimensión Temporal: El dataset cuenta con una granularidad mensual, cubriendo un periodo que permite analizar tendencias y estacionalidades.

Metodología y Procesamiento de Datos

Preparando los Datos para la Predicción

1. Segmentación Estratégica: Para asegurar un análisis preciso, los clientes fueron agrupados en quintiles según su nivel de acreditación de haberes. Esto permite trabajar con grupos homogéneos en términos de ingreso y comportamiento financiero.
2. Análisis de Correlación: Se estudió la relación entre las variables de consumo y los saldos dentro de cada segmento para validar las hipótesis iniciales.
3. Ingeniería de Características: La variable objetivo (saldo_promedio_vista) y otras variables monetarias fueron deflactadas con el IPC para eliminar el efecto de la inflación.
4. Preparación Final: Los datos de cada segmento fueron escalados (normalizados) y transformados en secuencias de entrada-salida para alimentar los modelos de Deep Learning.

Arquitecturas Implementadas

Los Contendientes: LSTM vs. Transformer

Modelo LSTM (Long Short-Term Memory):

Red Neuronal Recurrente diseñada específicamente para recordar patrones a largo plazo.

Ideal para capturar tendencias y estacionalidades en series temporales como los saldos bancarios.

Nuestra implementación usó una capa LSTM seguida de una capa lineal para el pronóstico.

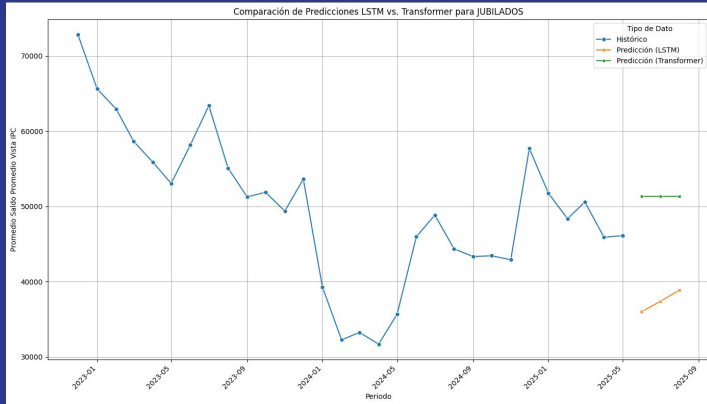
Modelo Transformer:

Arquitectura basada en mecanismos de auto-atención, que le permiten sopesar la importancia de cada punto de datos en la secuencia.

Se implementó un TransformerEncoder junto con un Positional Encoding para que el modelo entienda el orden de los datos.

Resultados de la Predicción de Saldos

Predicciones LSTM vs Transformer Por Segmento Comercial



Resultados de la Predicción de Saldos

Los Contendientes: LSTM vs. Transformer



Análisis Visual:

El modelo LSTM genera un pronóstico dinámico, que logra capturar parte de la volatilidad histórica de los saldos, haciéndolo más realista.

En contraste, el Transformer produce una predicción prácticamente constante, fallando en capturar las fluctuaciones naturales y la tendencia de los datos.

Métricas de Error (RMSE y MAE) para la predicción del saldo:

Cuadro Comparativo de RMSE y MAE Promediados (en Escala Original):

	Model	Average_RMSE_Original	Average_MAE_Original	
0	LSTM	156449.21	64347.91	
1	Transformer	156763.59	67851.97	

Conclusión

¿Qué Modelo Predice Mejor los Saldos?

Modelo Ganador: Para la tarea de pronosticar los saldos en cuentas vista, el modelo LSTM es superior.

Razones:

- Precisión Cuantitativa: Obtuvo un menor Error Cuadrático Medio (RMSE) y Error Absoluto Medio (MAE).
- Calidad del Pronóstico: Genera predicciones realistas y dinámicas que se alinean mucho mejor con el comportamiento de una serie temporal financiera.
- Utilidad Práctica: La predicción constante del Transformer, aunque estable, es poco útil porque no identifica tendencias, cambios o la volatilidad inherente al mercado.

Conclusión Final: Se validó que las variables económicas permiten predecir los saldos futuros, y que la arquitectura LSTM es la herramienta más adecuada para esta tarea específica gracias a su capacidad para modelar la dinámica de los datos.