

Análisis de series de tiempo financieras implementando métodos de TDA

Pablo Pérez Sanadoval

Instituto Tecnológico y de Estudios Superiores de Monterrey
Estancia de investigación (Gpo 572)
Docente: Jonathan Montalvo Urquiza

02 de diciembre de 2025

Contexto y motivación

- Pronóstico de series de tiempo financieras: problema clásico, pero difícil por ruido, no estacionariedad y cambios de régimen.
- Modelos modernos de aprendizaje profundo han mejorado el modelado de relaciones no lineales, pero suelen enfocarse en patrones locales.
- Topological Data Analysis (TDA) ofrece descriptores globales de la forma de la serie, potencialmente complementarios a indicadores clásicos.
- Idea central: combinar TDA con N-BEATSX para mejorar el pronóstico de log-retornos de acciones en un panel de mercado realista.

Objetivo y planteamiento del estudio

- Objetivo principal:
 - Cuantificar la contribución marginal de características de TDA como variables exógenas en el pronóstico de log-retornos diarios escalados.
- Enfoque:
 - Comparar un modelo N-BEATSX base (sin TDA) contra un N-BEATSX con descriptores topológicos agregados como exógenas históricas.
 - Mantener idéntica arquitectura, panel, definición de objetivo y variables exógenas no topológicas.
- Desarrollo en dos etapas:
 - Estudio piloto con implementación manual de N-BEATS y preprocessamiento agresivo.
 - Diseño experimental rediseñado y más robusto usando neuralforecast y un objetivo más simple.

Antecedentes técnicos clave

■ Topological Data Analysis (TDA):

- Usa embeddings de retrasos, complejos simpliciales y homología persistente para capturar componentes, bucles y cavidades.
- Los diagramas de persistencia se resumen en vectores de características (por ejemplo, amplitud, entropía, número de puntos).

■ TDA en finanzas:

- Se ha aplicado para detectar cambios de régimen, analizar redes de correlación y construir características para pronóstico.
- Trabajos previos muestran que las características topológicas pueden capturar complejidad de mercado no reflejada en indicadores estándar.

■ N-BEATSX:

- Arquitectura de pronóstico basada en bloques totalmente conectados con soporte para paneles y variables exógenas.
- Implementación pública en neuralforecast, adecuada para experimentos a gran escala.

Datos y preprocesamiento

- Panel con diez acciones de gran capitalización:
 - AAPL, MSFT, GOOGL, AMZN, NVDA, META, TSM, LLY, BRK-B, NVO.
- Periodo: datos diarios de mercado desde 2019 hasta 2025 (frecuencia de días hábiles).
- Construcción del panel en formato largo:
 - Columnas: `unique_id`, `ds`, `Y` más características exógenas.
- Variable objetivo:
 - Log-retornos diarios a partir de precios de cierre ajustados.
 - Escalado MinMax para obtener la serie `logret_scaled`, usada como objetivo en todos los modelos.
- Las ventanas deslizantes para TDA generan valores faltantes al inicio; se ajusta la fecha de inicio efectiva del panel.

Variables exógenas no topológicas

- **Exógenas futuras de calendario (futr_exog):**
 - dow, dom, woy, month, qtr.
 - Indicadores de fin de periodo: eom, eoq, eoy.
 - Información determinista conocida para todo el horizonte de pronóstico.
- **Exógenas históricas de estado de mercado (hist_exog):**
 - Índice de volatilidad implícita: vix.
 - Volatilidad realizada a 5, 10 y 21 días: rv_5, rv_10, rv_21.
- **Sin exógenas estáticas:**
 - stat_exog = None en todos los experimentos.
- Estas variables se mantienen idénticas entre el modelo base y el modelo con TDA, para aislar el efecto de las características topológicas.

Construcción de características TDA

- Flujo basado en trabajos previos, adaptado al contexto de acciones:
 - Uso de la serie escalada `logret_scaled` para cada activo.
- **Embedding de retrasos (Takens):**
 - Dimensión de embedding igual a 4 y retraso igual a 4, generando una nube de puntos en espacio de dimensión 4.
- **Homología persistente en dimensión uno:**
 - Construcción de un complejo de Vietoris–Rips sobre la nube de puntos.
 - Cálculo de diagramas de persistencia que capturan bucles en la trayectoria.
- **Resumen por ventanas deslizantes:**
 - Ventana de 21 días hábiles sobre la serie embebida.
 - Descriptores agregados: `tda_amplitude_h1_w21` y `tda_n_points_h1_w21`.
 - Ambas columnas se incorporan como exógenas históricas (`hist_exog`) en N-BEATSX.

Metodología de modelado

■ Formulación del problema:

- Pronóstico de log-retornos escalados a cinco días hábiles.
- Enfoque de panel: un único modelo N-BEATSX entrenado sobre las diez acciones.

■ Modelos comparados:

- Modelo base: N-BEATSX sin características de TDA.
- Modelo con TDA: misma configuración, más amplitud y número de puntos topológicos como exógenas históricas.

■ Diseño experimental:

- División temporal en entrenamiento y prueba respetando el orden del tiempo.
- Esquema de backtesting con cuatro ventanas de origen rodante, cada una con horizonte de cinco días hábiles.
- Ambas versiones del modelo se entrena y evalúan con las mismas divisiones y configuración.
- El estudio piloto previo mostró que un preprocesamiento demasiado agresivo y una implementación manual de N-BEATS degradaban el desempeño, lo que motivó la migración a neuralforecast y a un flujo más simple.

Resultados globales

■ Comparación de métricas agregadas en el panel:

- Reducción del error porcentual absoluto medio (MAPE) de 144.73 a 138.16 al incorporar TDA.
- Reducción del error absoluto medio (MAE) de 0.764 a 0.754.
- Disminución del error absoluto medio escalado (MASE) de 0.766 a 0.752.

■ Interpretación:

- El modelo con TDA mejora de forma consistente los errores absolutos y relativos, aunque el impacto en errores grandes es limitado.

■ Análisis de diferencias de error absoluto:

- Se define la diferencia de errores absolutos entre modelo con TDA y modelo base.
- En 64 por ciento de los puntos, el modelo con TDA presenta menor error absoluto que el modelo sin TDA.
- Las medias y medianas de estas diferencias son negativas, lo que indica una mejora típica moderada a favor de TDA.

Resultados por activo

■ Heterogeneidad entre acciones:

- En siete de los diez activos (AAPL, BRK-B, GOOGL, LLY, META, NVDA y TSM), la proporción de puntos donde el modelo con TDA tiene menor error absoluto es mayor a la mitad.
- TSM es el caso extremo: en todas las observaciones del horizonte de pronóstico, TDA mejora el error absoluto.
- AAPL y NVDA también muestran reducciones de error relativamente marcadas.

■ Acciones donde TDA no ayuda:

- En AMZN, MSFT y NVO la proporción de puntos con mejor desempeño corresponde al modelo sin TDA.
- En estos casos, las medias de las diferencias de error absoluto favorecen ligeramente al modelo base.

■ Distribución de diferencias:

- La mayoría de las diferencias de error absoluto se encuentran en un rango acotado y concentrado alrededor de valores negativos moderados.
- El efecto de TDA no es drástico punto a punto, pero sí introduce una mejora sistemática a nivel global.

Conclusiones y trabajo futuro

■ Conclusiones principales:

- Las características topológicas basadas en amplitud y número de puntos persistentes en diagramas de homología de dimensión uno aportan información útil al pronóstico de log-retornos escalados.
- El modelo con TDA supera de forma consistente al modelo base en métricas de error absoluto típicas, aunque las mejoras son moderadas.
- La estructura topológica implícita en los retornos parece contener señal que no es capturada por transformaciones o exógenas tradicionales.

■ Limitaciones del enfoque actual:

- N-BEATSX es un modelo de regresión para niveles futuros, no está alineado directamente con el objetivo práctico de trading.
- Para decisiones operativas, interesa más la dirección del movimiento que el nivel exacto.