



Pontificia Universidad
JAVERIANA
Cali

PREDICCIÓN DE LA TASA DE CAMBIO EUR/USD: COMPARACIÓN DE
MODELOS DE DEEP LEARNING CON DIVERSOS TIPOS Y CANTIDADES DE
DATOS DE ENTRENAMIENTO.

Harvey Bastidas

Andrés Caicedo

Cristian Sarmiento

Proyecto Aplicado para optar al título de

Magister en Ciencia de Datos

Directora:

Prof. María Constanza Pabón

FACULTAD DE INGENIERÍA Y CIENCIAS

MAESTRÍA EN CIENCIA DE DATOS

SANTIAGO DE CALI, JUNIO 2025

RESUMEN

La predicción de la tasa de cambio del par EUR/USD es un desafío significativo debido a la naturaleza no lineal de las series temporales financieras, las técnicas de Deep Learning son capaces de detectar patrones no lineales y son usadas para este tipo de predicciones, pero su desempeño varía dependiendo de los datos de entrenamiento usados. Este proyecto se enfoca en evaluar cómo la cantidad y el tipo de datos de entrenamiento, incluyendo datos sintéticos, afectan la precisión y la robustez de modelos predictivos de Deep Learning. Este proyecto aborda la necesidad de seleccionar adecuadamente la configuración de modelos y datos para mejorar las predicciones a corto y largo plazo de tres modelos de Deep Learning: Redes Neuronales Convolucionales (CNN), Redes de Memoria a Largo Plazo (LSTM) y Transformadores. Se midió, usando métricas de precisión y robustez, el desempeño de los modelos entrenados con distintas cantidades de datos de entrenamiento y la inclusión de distintos tipos de datos como indicadores técnicos y fundamentales. Además, se evaluó la efectividad de los datos sintéticos de entrenamiento en el desempeño de los modelos. El objetivo principal es determinar cómo diferentes configuraciones de modelos y datos afectan la precisión y robustez de las predicciones de la tasa de cambio EUR/USD. Con esto se espera crear un conjunto de resultados que permitan seleccionar la mejor configuración de modelos y datos a usar para el entrenamiento de predictores, lo que puede ser una herramienta académica y puede usarse en la toma de decisiones en el mercado de divisas, ya que la predicción de tasas es crucial para elaborar estrategias de trading y desarrollar herramientas analíticas avanzadas para el sector financiero.

Contenido

Resumen.....	2
Índice de figuras	5
Índice de tablas	7
1. Introducción	9
2. Definición del problema	10
3. Objetivos	11
3.1. Objetivo General.....	11
3.2. Objetivos Específicos	11
4. Marco Teórico.....	12
4.1. Random Forest	12
4.2. Deep Learning.....	13
4.3. Otras arquitecturas de redes neuronales aplicadas a la predicción EUR/USD.....	14
4.4. Teoría de la Información aplicada a la selección de periodicidad	14
4.5. Uso de Datos Sintéticos	16
4.6. Métodos de descomposición y análisis espectral de series de tiempo.....	18
4.7. Preprocesamiento y Manejo de Datos	18
4.8. Métricas de Precisión	19
4.9. Métricas de Robustez	20
4.10. Conceptos de trading relevantes en el proyecto	20
5. Antecedentes	22
6. Metodología.....	24
6.1. Datasets Descargados.....	24
6.2. Arquitectura General del Sistema.....	25
6.3. Selección de Periodicidad para predicciones	26
6.4. Horizonte Predictivo	27
6.5. Determinación de Horizonte Temporal de Largo Plazo	27
6.6. Determinación de Horizonte Temporal de corto Plazo	28
6.7. Estrategia heurística de trading basada en predicciones de corto y largo plazo.....	28
6.8. Experimentos para cumplir objetivos.....	29
7. Exploración de los Datos	32
7.1. Descomposición con STL	35

8.	Ingeniería de características	37
9.	Preparación de los Datos.....	38
9.1.	Preprocesamiento de información	38
10.	Preparación de Datasets específicos para cada Modelo Predictivo	39
10.1.	Predictores con ANN, LSTM Y Conv1D.....	39
10.2.	Predictores con Transformers.....	39
11.	Modelado.....	40
11.1.	Modelos de Deep Learning e hiperparámetros.....	40
11.2.	Modelo de artificial Neural Networks (ANN) con capas densas.....	41
11.3.	Modelo de red Neural Convolutacional Unidimensional (CNN)	43
11.4.	Modelo LSTM.....	44
11.5.	Modelo basado en Tranformers	45
12.	Desarrollo de experimentos.....	48
12.1.	Experimentos de variación de la cantidad de registros de entrenamiento.....	48
12.2.	Experimento de variación de número y tipo de características de los datos de entrenamiento.....	57
12.3.	Experimento de Uso de Extractor de Características con autoencoder	81
12.4.	Experimentos de Uso de datos sintéticos para entrenamiento	90
13.	Análisis de Resultados de Simulación de Trading	100
13.1.	Parámetros Optimizados de la Estrategia de Trading Heurística	100
13.2.	Variación de cantidad de registros de entrenamiento	101
13.3.	Variación del tipo de características	105
13.4.	Uso de Extractores de Características	110
13.5.	Relación entre precisión del modelo y rendimiento de trading.....	112
14.	Conclusiones	118
15.	Trabajo futuro	120
16.	Referencias Bibliográficas	121

ÍNDICE DE FIGURAS

<i>Figura 1 - Serie de tiempo con valores de cierre a 24 horas</i>	32
<i>Figura 2 - Serie de tiempo con valores de cierre a una hora</i>	33
<i>Figura 3 - Descomposición STL de periodicidad largo plazo</i>	35
<i>Figura 4 - Distribución de la serie de cierre por día de la semana</i>	36
<i>Figura 5 - Distribución de la serie de cierre por mes del año en formato diario</i>	36
<i>Figura 6 – Modelo de capas densas ANN con Parallel Branch Feature Processing y Direct Multi-Output (Fragmento de plot generado con keras.utils.plot_model)</i>	42
<i>Figura 7 – Modelo CNN con Direct Multi-Output (Fragmento de plot generado con keras.utils.plot_model)</i>	43
<i>Figura 8 – Modelo Bidir-LSTM con Direct Multi-Output (Fragmento de plot generado con keras.utils.plot_model)</i>	45
<i>Figura 9 – Extractor de características de Modelo de Transformers con Direct Multi-Output (Fragmento de plot generado con keras.utils.plot_model)</i>	47
<i>Figura 10 - Desempeño de MAE al aumentar la cantidad de datos - corto plazo</i>	52
<i>Figura 11 - Predicción para CNN con 25.200 steps con periodos de corto plazo</i>	53
<i>Figura 12 - Desempeño de MAE al aumentar la cantidad de datos en el horizonte de largo plazo</i>	56
<i>Figura 13 - Predicción para ANN con 25.200 steps con periodos de largo plazo</i>	57
<i>Figura 14 - Comparación entre valores reales y predichos por el modelo CNN utilizando indicadores técnicos - corto plazo</i>	62
<i>Figura 15 - Comparación entre valores reales y predichos por el modelo ANN utilizando indicadores técnicos – largo plazo</i>	62
<i>Figura 16 - Comparación entre valores reales y predichos por el modelo CNN adicionando datos estacionales – corto plazo</i>	65
<i>Figura 17 - Comparación entre valores reales y predichos por el modelo ANN adicionando datos estacionales – largo plazo</i>	66
<i>Figura 18 - Comparación entre valores reales y predichos por el modelo CNN adicionando datos de alta frecuencia – corto plazo</i>	68
<i>Figura 19 - Comparación entre valores reales y predichos por el modelo ANN adicionando datos de alta frecuencia – largo plazo</i>	69
<i>Figura 20 - Comparación entre valores reales y predichos por el modelo CNN adicionando indicadores fundamentales - corto plazo</i>	72
<i>Figura 21 - Comparación entre valores reales y predichos por el modelo ANN adicionando indicadores fundamentales largo plazo Adición de todas las características al modelo</i>	72
<i>Figura 22 – Cabezales predictivos usados en experimentos de sección 12.10 y 13 (Fragmento de model plot generado con Keras)</i>	73
<i>Figura 23 - Comparación entre valores reales y predichos por el modelo ANN adicionando todas las características - corto plazo</i>	75
<i>Figura 24 - Comparación entre valores reales y predichos por el modelo ANN adicionando todas las características largo plazo</i>	75
<i>Figura 25 - Comparación real vs predicho (corto plazo) usando la mejor combinación (Extractor ANN + Modelo Predictivo CNN)</i>	83
<i>Figura 26 - Comparación real vs predicho (largo plazo) usando la mejor combinación (actualmente Extractor ANN + Modelo Predictivo CNN)</i>	84
<i>Figura 27 - Modelo del Autoencoder usando capas Conv1D, generado usando Keras</i>	85
<i>Figura 28 – Fragmento del Modelo del Autoencoder usando capas Dense (ANN) en paralelo para cada característica de entrada, plot generado usando Keras</i>	86

Figura 29 - Comparación real vs predicho (corto plazo) usando la mejor combinación con extractor pre-entrenado..... 89

Figura 30 - Comparación real vs predicho (largo plazo) usando la mejor combinación con extractor pre-entrenado..... 90

Figura 31 – Plot de modelo de VAE 91

Figura 32 - Balance de Trading usando predicciones con LSTM entrenado con 3150 registros. 103

Figura 33 - Balance de Trading usando predicciones con LSTM entrenado con 12600 registros. 104

Figura 34 - Balance de Trading de baja frecuencia usando predicciones con ANN entrenado con 25200 registros. 108

Figura 35 - Balance de Trading de alta frecuencia usando predicciones con LSTM entrenado con 25200 registros. 109

Figura 36 - Balance de Trading de Baja frecuencia usando predicciones con extractor CNN. 111

Figura 37 - Balance de Trading de Alta frecuencia usando predicciones con extractor CNN. 112

Figura 38 - Profit vs. MAE de Test (Largo Plazo) - Alta Frecuencia. 113

Figura 39 - Profit vs. MAE de Test (Largo Plazo) - Baja Frecuencia. 114

Figura 40 - Balance plot de trading con predicciones Ideales - Alta Frecuencia..... 115

Figura 41 - Balance plot de trading con predicciones Ideales - Alta Frecuencia..... 115

Figura 42 - Ratio de Sharpe vs. MAE de Test (Largo Plazo) - Alta Frecuencia..... 116

Figura 43 - Ratio de Sharpe vs. MAE de Test (Largo Plazo) - Baja Frecuencia. 117

ÍNDICE DE TABLAS

<i>Tabla 1 - Estadísticas para Selección de Periodicidades de Corto y Largo Plazo.....</i>	<i>26</i>
<i>Tabla 2 - Consolidado de experimentos para cumplir objetivos</i>	<i>31</i>
<i>Tabla 3 - Análisis exploratorio básico de los datos</i>	<i>32</i>
<i>Tabla 4 - Resultados MAE de experimentos de variación de número de registros de entrenamiento para corto plazo.....</i>	<i>49</i>
<i>Tabla 5 - Resultados R2 de experimentos de variación de número de registros de entrenamiento para corto plazo.....</i>	<i>51</i>
<i>Tabla 6 - Resultados MAE de experimentos de variación de número de registros de entrenamiento para largo plazo.....</i>	<i>54</i>
<i>Tabla 7 - Resultados R2 de experimentos de variación de número de registros de entrenamiento para largo plazo.....</i>	<i>55</i>
<i>Tabla 8 - Resultados MAE de experimentos adición de indicadores técnicos</i>	<i>59</i>
<i>Tabla 9 - Resultados R2 de experimentos adición de indicadores técnicos</i>	<i>60</i>
<i>Tabla 10 - Resultados SNR de experimentos adición de indicadores técnicos</i>	<i>61</i>
<i>Tabla 11 - Resultados MAE de experimentos adición de datos estacionales.....</i>	<i>63</i>
<i>Tabla 12 - Resultados R2 de experimentos adición de datos estacionales</i>	<i>64</i>
<i>Tabla 13 - Resultados SNR de experimentos adición de datos estacionales.....</i>	<i>64</i>
<i>Tabla 14 - Resultados MAE de experimentos adición de datos alta frecuencia.....</i>	<i>67</i>
<i>Tabla 15 - Resultados R2 de experimentos adición de datos alta frecuencia</i>	<i>67</i>
<i>Tabla 16 - Resultados SNR de experimentos adición de datos alta frecuencia.....</i>	<i>68</i>
<i>Tabla 17 - Resultados MAE de experimentos adición de indicadores fundamentales.....</i>	<i>70</i>
<i>Tabla 18 - Resultados R2 de experimentos adición de indicadores fundamentales</i>	<i>70</i>
<i>Tabla 19 - Resultados SNR de experimentos adición de indicadores fundamentales.....</i>	<i>71</i>
<i>Tabla 20 - Resultados MAE de experimentos adición de todas las características</i>	<i>73</i>
<i>Tabla 21 - Resultados R2 de experimentos adición de todas las características</i>	<i>74</i>
<i>Tabla 22 - Resultados SNR de experimentos adición de todas las características</i>	<i>74</i>
<i>Tabla 23 - Comparativa MAE Validación - Variación Cantidad de Datos vs Experimento con Indicadores Técnicos.....</i>	<i>76</i>
<i>Tabla 24 - Comparativa MAE Validación - Experimento de Variación de Cantidad de Datos vs Experimento con Datos Estacionales</i>	<i>77</i>
<i>Tabla 25 - Comparativa MAE Validación - Variación Cantidad de Datos vs Experimento con adición de datos de Alta Frecuencia.....</i>	<i>77</i>
<i>Tabla 26 - Comparativa MAE Validación - Experimento de Variación de Cantidad de Datos. vs Experimento de Adición de Datos Fundamentales.....</i>	<i>78</i>
<i>Tabla 27 - Comparativa de resultados MAE base vs MAE todas las características.....</i>	<i>79</i>
<i>Tabla 28 - Métricas de Desempeño (MAE) - Extractores sin Pre-entrenamiento</i>	<i>82</i>
<i>Tabla 29 - Métricas de Desempeño (R²) - Extractores sin Pre-entrenamiento.....</i>	<i>82</i>
<i>Tabla 30 - Métricas de Desempeño (SNR) - Extractores sin Pre-entrenamiento</i>	<i>83</i>
<i>Tabla 31 - MAE de Reconstrucción - Entrenamiento de Autoencoders.....</i>	<i>87</i>
<i>Tabla 32 - Métricas (MAE) - Extractores Pre-entrenados</i>	<i>88</i>
<i>Tabla 33 - Métricas (R²) - Extractores Pre-entrenados</i>	<i>88</i>
<i>Tabla 34 - Métricas (SNR) - Extractores Pre-entrenados</i>	<i>88</i>
<i>Tabla 35 - Resultados MAE de experimento de uso de datos sintéticos para corto plazo</i>	<i>92</i>
<i>Tabla 36 - Resultados R2 de experimento de uso de datos sintéticos para corto plazo.....</i>	<i>93</i>
<i>Tabla 37 - Resultados SNR de experimento de uso de datos sintéticos para corto plazo</i>	<i>93</i>
<i>Tabla 38 - Resultados MAE de experimento de uso de datos sintéticos para largo plazo</i>	<i>94</i>

<i>Tabla 39 - Resultados R2 de experimento de uso de datos sintéticos para largo plazo.....</i>	<i>94</i>
<i>Tabla 40 - Resultados SNR de experimento de uso de datos sintéticos para corto plazo</i>	<i>95</i>
<i>Tabla 41 - Resultados MAE de experimento de uso de datos sintéticos para corto plazo</i>	<i>96</i>
<i>Tabla 42 - Resultados R2 de experimento de uso de datos sintéticos para corto plazo.....</i>	<i>97</i>
<i>Tabla 43 - Resultados SNR de experimento de uso de datos sintéticos para corto plazo</i>	<i>97</i>
<i>Tabla 44 - Resultados MAE de experimento de uso de datos sintéticos para largo plazo</i>	<i>98</i>
<i>Tabla 45 - Resultados R2 de experimento de uso de datos sintéticos para largo plazo.....</i>	<i>98</i>
<i>Tabla 46 - Resultados SNR de experimento de uso de datos sintéticos para corto plazo</i>	<i>99</i>
<i>Tabla 47 - Resultados de Simulación (Baja Frecuencia) - Experimento: Variación de Cantidad de Registros</i>	<i>102</i>
<i>Tabla 48 - Resultados de Simulación (Alta Frecuencia) - Experimento: Variación de Cantidad de Registros</i>	<i>102</i>
<i>Tabla 49 - Resultados de Simulación (Baja Frecuencia) - Adición de Indicadores Técnicos</i>	<i>105</i>
<i>Tabla 50 - Resultados de Simulación (Alta Frecuencia) - Adición de Indicadores Técnicos</i>	<i>105</i>
<i>Tabla 51 - Resultados de Simulación (Baja Frecuencia) - Adición de Datos Estacionales.....</i>	<i>106</i>
<i>Tabla 52 - Resultados de Simulación (Alta Frecuencia) - Adición de Datos Estacionales.....</i>	<i>106</i>
<i>Tabla 53 - Resultados de Simulación (Baja Frecuencia) - Adición de Datos de Alta Frecuencia</i>	<i>106</i>
<i>Tabla 54 - Resultados de Simulación (Alta Frecuencia) - Adición de Datos de Alta Frecuencia</i>	<i>106</i>
<i>Tabla 55 - Resultados de Simulación (Baja Frecuencia) - Adición de Indicadores Fundamentales.....</i>	<i>107</i>
<i>Tabla 56 - Resultados de Simulación (Alta Frecuencia) - Adición de Indicadores Fundamentales.....</i>	<i>107</i>
<i>Tabla 57 - Resultados de Simulación (Baja Frecuencia) - Adición de Todas las Características.....</i>	<i>108</i>
<i>Tabla 58 - Resultados de Simulación (Alta Frecuencia) - Adición de Todas las Características.....</i>	<i>108</i>
<i>Tabla 59 - Resultados de Simulación (Baja Frecuencia) - Extractores sin Pre-entrenamiento + Predictor Fijo</i>	<i>110</i>
<i>Tabla 60 - Resultados de Simulación (Alta Frecuencia) - Extractores sin Pre-entrenamiento + Predictor Fijo</i>	<i>110</i>
<i>Tabla 61 - Resultados de Simulación (Baja Frecuencia) - Extractores Pre-entrenados + Predictor Fijo</i>	<i>111</i>
<i>Tabla 62 - Resultados de Simulación (Alta Frecuencia) - Extractores Pre-entrenados + Predictor Fijo.....</i>	<i>111</i>

1. INTRODUCCIÓN

Las tasas de cambio de las divisas están influenciadas por diversos factores económicos, políticos y sociales lo que genera comportamientos no-lineales que son difíciles de predecir. Este problema se ve exacerbado por la limitada capacidad de los modelos estadísticos tradicionales para capturar estas dinámicas complejas, lo que hace necesaria la adopción de técnicas avanzadas de Deep Learning. Pero los modelos usados en estas técnicas requieren ser ajustados con datos de entrenamiento, de los cuales existe una gran variedad de posibles selecciones de cantidad y tipo de datos a usar, que pueden influir en el desempeño de los predictores y se requiere un criterio de selección de estos factores.

La problemática de la selección de los datos a usar en el entrenamiento de modelos predictivos de Deep Learning es importante, ya que una predicción precisa de la tasa de cambio EUR/USD es fundamental para diversos actores económicos, incluyendo exportadores, importadores, inversores y entidades financieras. Estas predicciones son cruciales para elaborar estrategias de trading eficaces y desarrollar herramientas analíticas avanzadas que faciliten la toma de decisiones informadas y la gestión de riesgos cambiarios.

Este proyecto se propuso evaluar cómo la cantidad y el tipo de datos de entrenamiento, incluyendo datos sintéticos, afectan la precisión y robustez de tres modelos predictivos de Deep Learning, específicamente Redes Neuronales Convolucionales (CNN), Redes de Memoria a Largo Plazo (LSTM) y Transformadores. Se investigó el impacto de distintas configuraciones de modelos y datos, abarcando tres cantidades de datos de entrenamiento, por ejemplo: 10,000, 100,000 y 1,000,000 registros con periodicidad de 5 minutos, aunque la periodicidad para los experimentos se seleccionó durante el desarrollo del proyecto. También se incluyeron indicadores técnicos, fundamentales y datos sintéticos a los datos de entrenamiento.

Se llevó a cabo un análisis utilizando métricas de precisión y robustez para evaluar el desempeño de los tres modelos. El proyecto también incluyó la generación de datos sintéticos con el decodificador, mientras el codificador fue usado como extractor de características para reducción de ruido y dimensionalidad.

El resultado obtenido son las comparaciones de precisión y robustez entre los modelos predictivos, para predicciones tanto en el largo como el corto plazo, al ser entrenados con distintos tipos y cantidades de datos, lo que permite seleccionar de manera informada, la configuración de modelos y datos a usar para predicciones. Esto puede ser útil tanto para proyectos académicos como para uso en la industria financiera.

2. DEFINICIÓN DEL PROBLEMA

La predicción de la tasa de cambio y volatilidad del EUR/USD representa un desafío significativo debido a la naturaleza no lineal de las series temporales financieras. Los modelos de Deep Learning son una herramienta útil para la detección de patrones no lineales que pueden mejorar las predicciones respecto a modelos estadísticos tradicionales. Ya que en el ámbito financiero es útil conocer tanto las predicciones a largo, como corto plazo y dado que existe una variedad de modelos de Deep Learning y tipos de datos de entrenamiento, incluyendo datos sintéticos, se presenta una dificultad para seleccionar la configuración correcta de modelos, tipos de datos y cantidad de datos a usar para el entrenamiento de los modelos predictivos.

La documentación existente en el dominio del Deep Learning aplicado al mercado financiero, se enfoca en medición de desempeño de modelos predictivos y comparación de desempeño entre modelos, pero muy poca de esta información se centra en la selección de los datos a usar para predicción. Dado que la cantidad y tipo de datos es un factor importante para tener en cuenta al entrenar un predictor, tener información del desempeño de los modelos de Deep Learning más importantes, con diferentes configuraciones de tipos y cantidades de datos, puede ser útil para tener una buena predicción. Tener una predicción confiable de la tasa de cambio EUR/USD es fundamental para diversos actores económicos, incluyendo exportadores, importadores, inversores y entidades financieras. Una mejor comprensión de estos movimientos permite optimizar estrategias de inversión, gestionar riesgos cambiarios y tomar decisiones financieras informadas [1].

Los síntomas del problema se manifiestan en predicciones que no logran capturar la tendencia, resultando en señales falsas o retrasadas que pueden llevar a tomar malas decisiones financieras. Estos problemas se atribuyen a varios factores, incluyendo la insuficiencia de datos históricos relevantes, el sobreajuste de modelos en datos pasados que no reflejan condiciones futuras y la falta de capacidad para incorporar adecuadamente nuevos indicadores económicos o geopolíticos en tiempo real [2]. También, la limitada dimensionalidad y cantidad de datos utilizados en algunos casos no capturan la verdadera complejidad del mercado, o no logran ajustar los modelos suficientemente, limitando así su efectividad. Al no existir en el conocimiento de los autores, una comparación sistemática de los principales modelos de Deep Learning entrenados con diferentes tipos de datos, incluyendo datos sintéticos, para diferentes horizontes temporales, se puede usar una configuración no-óptima de datos para un modelo específico.

Desde la perspectiva de la ciencia de datos, el problema radica en cómo seleccionar el tipo y cantidad de datos, incluyendo datos sintéticos, para usar en el entrenamiento de los principales modelos predictivos de Deep Learning, de manera que se aproveche la capacidad predictiva a largo y corto plazo, de mejor manera que usando unos tipos y cantidades de datos seleccionados sin un criterio definido. Medir tanto la precisión como la robustez de las predicciones de distintas configuraciones de datos y modelos, permitirá tener un criterio de selección que puede ser útil para un científico de datos que trabaja en este dominio. Tanto las técnicas de Deep Learning, como el uso de datos sintéticos, son áreas de investigación en curso [3][4].

3. OBJETIVOS

En esta sección se muestra tanto el objetivo general como los objetivos específicos.

3.1. OBJETIVO GENERAL

Determinar el impacto de la cantidad y tipo de datos de entrenamiento, incluyendo datos sintéticos, en la precisión y robustez de 3 modelos predictivos (CNN, redes de LSTM y Transformers) para la tasa de cambio EUR/USD para el corto y largo plazo con el fin de tener un criterio de selección de modelos y datos para obtener la predicción más consistente con el comportamiento de las series.

3.2. OBJETIVOS ESPECÍFICOS

- Identificar una fuente confiable de datos que contenga la cantidad de datos requeridos para cumplir los demás objetivos, descargarlos y preprocesarlos para su utilización.
- Medir cómo el incremento de la cantidad de los datos de entrenamiento influye en la precisión y robustez de la predicción de la tasa EUR/USD con los modelos seleccionados de Deep Learning.
- Medir cómo el incremento de la dimensionalidad de los datos de entrenamiento, por medio de la adición de diferentes tipos de datos (indicadores, técnicos, fundamentales y otras tasas), influye en la precisión y robustez de la predicción de la tasa EUR/USD con los modelos seleccionados.
- Medir como la adición de datos sintéticos a los datos históricos de entrenamiento de los modelos seleccionados, influye en la precisión y robustez de la predicción de la tasa EUR/USD, respecto al uso únicamente de datos históricos en el entrenamiento de los modelos.
- Evaluar los modelos predictivos seleccionados para realizar predicciones en el corto (horas) y largo plazo (semanas).

4. MARCO TEÓRICO

En esta sección se revisan las principales teorías y conceptos que se usaron para desarrollar los objetivos propuestos.

4.1. RANDOM FOREST

Random Forest es un método de aprendizaje automático que consiste en la construcción de múltiples árboles de decisión a partir de muestras aleatorias (bootstrap) del conjunto de datos, lo que permite capturar relaciones no lineales y complejas entre las variables. Cada árbol del bosque se entrena de manera independiente y, posteriormente, se combinan sus resultados mediante un proceso de votación (para clasificación) o promediado (para regresión), lo que mejora la robustez y la precisión del modelo al reducir el riesgo de sobreajuste. Este enfoque basado en la agregación de múltiples predicciones individuales hace que Random Forest sea especialmente adecuado para manejar datos multivariados y de alta dimensionalidad.

En este proyecto de grado, este método se emplea como base comparativa para evaluar el desempeño de nuestros predictores de deep learning, aprovechando su capacidad para manejar conjuntos de datos multivariados y generar modelos robustos a partir de un entrenamiento. La metodología de Random Forest permite realizar una partición clara entre el conjunto de entrenamiento y el de validación, facilitando así una evaluación objetiva para una comparación con otros modelos de machine learning en escenarios complejos, como el actual, donde se comparó el desempeño de diferentes modelos en un dataset de validación, habiendo sido entrenados en un dataset de entrenamiento.

En contraste, el modelo ARIMA, aunque ampliamente utilizado en el análisis de series temporales, presenta limitaciones significativas para trabajar con datos multivariados. ARIMA se centra en la predicción punto a punto, donde la media móvil debe calcularse en cada instante a partir de los valores inmediatamente anteriores, lo cual lo hace poco adecuado para realizar predicciones a largo plazo, como los requeridos en la evaluación de modelos en el dataset de validación. En nuestro caso, utilizar ARIMA implicaría enfrentar la dificultad de entrenar un modelo con un conjunto de training y evaluar posteriormente en un dataset de validación extenso (un año de días hábiles, es decir, 6240 horas), lo cual no es factible dada la naturaleza secuencial y puntual de sus predicciones.

Por otro lado, Random Forest permite construir un modelo integral que se entrena de manera completa sobre el conjunto de entrenamiento disponible, y luego se evalúa en el dataset de validación, proporcionando una medida comparable con otros modelos, del rendimiento del modelo en un periodo prolongado. Esta ventaja lo convierte en una opción idónea para contrastar los resultados obtenidos con modelos de Deep Learning, facilitando la comparación de desempeño en un entorno multivariado. De esta forma, al utilizar Random Forest como referencia, aseguramos una evaluación más coherente y robusta de las distintas técnicas implementadas en el proyecto [5].

4.2. DEEP LEARNING

El desarrollo de modelos de Deep Learning, ha permitido avances significativos en la predicción de series temporales financieras respecto a los métodos tradicionales, abordando eficazmente las complejidades inherentes a estos datos por medio del uso de múltiples niveles de abstracción y detección de patrones temporales complejos. Los modelos de Deep Learning usados en este proyecto son:

- **Redes Neuronales Convolucionales (CNN):** Las CNN son efectivas para capturar patrones locales en los datos de series temporales, identificando patrones relevantes en múltiples escalas de tiempo, por medio de usar un diseño modular con múltiples capas que realizan la operación de convolución, es decir el producto punto de un conjunto de filtros por capa llamado “kernel de convolución”, sobre el vector de entrada, esta operación produce un mapa de características que sirve de entrada a la siguiente capa. Generalmente en una CNN las capas convolucionales están separadas por otras capas como capas de activación ReLU, capas de pooling, capas de normalización, dropout, capas densas (ANN totalmente conectadas) e incluso capas de otros modelos de Deep Learning. Esta capacidad y flexibilidad, las hace ideales para el análisis de series temporales donde la precisión en la detección de patrones temporales es crucial, se pueden usar para detectar patrones locales en una o más dimensiones, las CNN 2D son usadas principalmente en procesamiento de imágenes y las CNN LARGO PLAZO para series de tiempo [10].
- **Redes Neuronales Recurrentes (RNN):** Las redes recurrentes, incluyendo RNN y Long-Short Time Memory Networks (**LSTM**), presentan buen desempeño en el campo del análisis de series temporales por su capacidad para procesar secuencias de datos con dependencias a largo plazo y detectar patrones temporales en diferentes escalas de tiempo. Algunas variantes como las Bidireccional LSTM, presentan mejor desempeño que las LSTM originales. Estas redes recuerdan información pasada durante períodos prolongados, esencial para predecir la volatilidad del mercado y otros indicadores financieros, pero su capacidad de detección de secuencias puede afectar cuando existen patrones sin una secuencia definida, que se dan principalmente en las predicciones a corto plazo [11].
- **Redes de Transformadores:** Los transformadores han mostrado un gran potencial en el análisis de series temporales financieras, especialmente a corto plazo, gracias a su arquitectura basada en mecanismos de atención, que consiste en asignar pesos a los diferentes filtros que se buscan en cada capa para producir el siguiente mapa de características, pero estos pesos (que simulan el mecanismo de atención humana) varían y se ajustan a medida que el modelo recibe nuevas entradas. Esto, les permite detectar patrones y dependencias temporales complejas sin las restricciones de secuencialidad impuestas por las RNN. Esta capacidad permite buen desempeño en las series temporales a corto plazo, aunque para predicciones a largo plazo su desempeño disminuye [12].
- **Positional Encoding:** Constituye un mecanismo para el procesamiento de secuencias temporales. A diferencia de modelos recurrentes, el Transformer procesa todos los elementos de la secuencia de entrada de manera concurrente, careciendo de una comprensión implícita del orden secuencial. El Positional Encoding introduce información explícita sobre la ubicación de cada elemento dentro de la secuencia directamente en su representación vectorial.

4.3. OTRAS ARQUITECTURAS DE REDES NEURONALES APLICADAS A LA PREDICCIÓN EUR/USD

Además de **ANN**, **CNN**, **LSTM** y **Transformer**, tres familias de redes han mostrado contribuciones sobresalientes en la literatura sobre forecasting del par EUR/USD:

- **GRU (Gated Recurrent Units)**. Su diseño compacto, con una sola puerta de actualización, reduce la cantidad de parámetros respecto al LSTM y acelera el entrenamiento sin perder capacidad de memoria. Altuntaş muestra que un modelo GRU mejora el MAE intradía frente a LSTM y XGBoost en la predicción del euro spot [28].
- **Bi-LSTM (Bidirectional LSTM)**. Procesa la serie en ambos sentidos, capturando dependencias pasadas y futuros latentes. Troiano y Restaino reportan una caída cercana al 8 % en MAE, comparado con un LSTM unidireccional, para velas de cuatro horas del EUR/USD [29].
- **Temporal Convolutional Network (TCN)**. Emplea convoluciones dilatadas con *skip connections* para abarcar ventanas largas sin recurrencias. Liang et al. recopilan varios estudios en los que la TCN reduce MAPE y acelera la inferencia frente a ANN tradicionales en datos de alta frecuencia de pares EUR/USD y GBP/USD [30].

4.4. TEORÍA DE LA INFORMACIÓN APLICADA A LA SELECCIÓN DE PERIODICIDAD

Para determinar la periodicidad (frecuencia) que aporta más información útil y menos ruido a los modelos de predicción, se evaluaron cinco métricas basadas en estadística y teoría de la información [27].

A continuación se definen y se describe su cálculo.

4.4.1. RELACIÓN SEÑAL-RUIDO (SNR)

La **SNR (Signal-to-Noise Ratio)** es la razón entre la potencia de la señal P_S y la potencia del ruido P_N :

$$SNR = \frac{P_S}{P_N}$$

- **Potencia de la señal P_S** : varianza de la serie de precios filtrada (o de los retornos).
- **Potencia del ruido P_N** : varianza de los residuos después de aplicar un modelo de suavizado (p. ej. Hodrick-Prescott) o un filtro de Kalman.

Un valor alto de SNR indica una serie con menor proporción de ruido, deseable para modelos predictivos estables.

4.4.2. COCIENTE RUIDO / RETORNOS (RR)

El **RR (Noise-to>Returns Ratio)** cuantifica cuánta variabilidad proviene del ruido respecto a la variabilidad total de los retornos σ_R :

$$RR = \frac{\sigma_{Ruido}}{\sigma_R}$$

- σ_{Ruido} : desviación estándar de los residuos.
- σ_R : desviación estándar de los retornos logarítmicos.

Cuanto menor sea RR, mejor relación señal/ruido presenta el conjunto de datos.

4.4.3. INFORMACIÓN ÚTIL LIBRE DE RUIDO (NR)

La métrica **NR** relaciona la información útil (libre de ruido) contenida en la serie y la calidad medida por RR:

$$NR = \frac{\text{Información libre de ruido}}{RR}$$

En la práctica se utilizó:

$$NR = \frac{P_S}{P_N} * \frac{1}{RR} = SNR * \frac{1}{RR}$$

Un valor alto de NR implica datos con gran contenido informativo y baja contaminación de ruido.

4.4.4. CAPACIDAD DE CANAL DE SHANNON–HARTLEY (C)

En un canal con ancho de banda BBB (Hz) y relación señal-ruido S/NS/NS/N, la capacidad máxima de transmisión de información viene dada en bits sobre segundo por:

$$C = B \log_2\left(1 + \frac{S}{N}\right)$$

Para series temporales financieras:

- $B = 1/\text{periodicidad}$ (p. ej. para datos de 1 minuto, $B = 1/60$ Hz).
- S/N es la SNR calculada en 4.3.1.

4.4.5. INFORMACIÓN LIBRE DE RUIDO EN BITS (I)

La información total libre de ruido disponible en un dataset de duración T (segundos) se obtiene multiplicando la capacidad C por dicho intervalo:

$$I = C * T$$

Este valor permite comparar objetivamente qué frecuencia (1 min, 15 min, 1 h, etc.) entrega más información útil para los modelos.

4.5. USO DE DATOS SINTÉTICOS

La generación de datos sintéticos es herramienta útil en el campo de la predicción de series temporales financieras, especialmente cuando se trata de superar limitaciones como la escasez de datos históricos o la necesidad de proteger la privacidad de los datos. Permiten a los investigadores y analistas aumentar sus conjuntos de datos, mejorar la robustez de los modelos y realizar pruebas extensivas sin comprometer datos sensibles. Además, pueden ayudar a explorar escenarios hipotéticos más allá de las experiencias históricas [13][14].

Es posible generar datos sintéticos tanto con métodos estadísticos como mediante modelos generativos, en este proyecto solo se usaron métodos generativos (VAE) para delimitar el alcance del proyecto y queda como trabajo futuro probar otros métodos.

Métodos Estadísticos para la Generación de Datos Sintéticos:

- **Simulación de Monte Carlo:** Este método implica la generación de múltiples muestras de variables aleatorias extraídas de una distribución obtenida de los datos, para simular una variedad de posibles resultados en procesos donde la intervención directa y experimentación real es imposible o impráctica. Se usa ampliamente en finanzas para evaluar y gestionar riesgos y para predecir cambios en los mercados [15].
- **Bootstrap:** El método Bootstrap permite la estimación de la distribución de un estadístico (como la media o la varianza) mediante el remuestreo repetido con reemplazo de los datos observados. Esto es particularmente útil para estimar la incertidumbre de las predicciones y para mejorar la robustez de los modelos financieros sin requerir nuevos datos externos [16].
- **Modelos de Cadena de Markov:** Estos modelos son efectivos para generar secuencias de datos que tienen dependencias complejas, basándose en la suposición de que el estado futuro depende solo del estado presente. Esto los hace adecuados para modelar decisiones financieras y cambios de mercado donde tales dependencias son prevalentes [17].

Modelos Generativos para obtener Datos Sintéticos:

- **Redes Generativas Antagónicas (GANs):** Las GANs son uno de los métodos más populares para la generación de datos sintéticos debido a su capacidad para producir muestras realistas. Una GAN se compone de dos redes neuronales que se entrenan de forma conjunta en un juego adversarial:
 - **Generador:** Recibe como entrada un vector de ruido (aleatorio) y produce datos sintéticos intentando imitar la distribución de los datos reales.
 - **Discriminador:** Evalúa tanto los datos reales como los generados, asignando una probabilidad de que cada muestra pertenezca al conjunto real.

El proceso de entrenamiento se orienta a que el generador mejore progresivamente, produciendo datos que el discriminador no pueda distinguir de los reales. A pesar de que

las GANs logran captar patrones locales y detalles finos de los datos, pueden presentar dificultades para reproducir con exactitud la distribución global a largo plazo [13].

- **Autoencoders Variacionales (VAEs):** Los VAEs son modelos basados en una arquitectura de codificador-decodificador que, además de aprender a reconstruir los datos de entrada, imponen una estructura probabilística en el espacio latente. Sus componentes principales son:
 - Encoder: Transforma los datos de entrada en una representación de menor dimensión, generando parámetros (media y varianza) que definen una distribución (normalmente gaussiana) en el espacio latente.
 - Capa de muestreo: A partir de la media y varianza obtenidas, se extraen muestras aleatorias que aseguran que la representación en el espacio latente cumpla con una distribución deseada.
 - Conditional Decoder: Reconstruye los datos originales a partir de la muestra extraída en el espacio latente, usando la fecha y los valores generados en el anterior tick de cada característica como entrada condicional para generación secuencial.

Esta estructura permite que los VAEs generen nuevos datos que preservan las propiedades estadísticas de los datos de entrenamiento. Son especialmente útiles para la extracción de características y la generación de muestras sintéticas coherentes [14].

- **Método Híbrido VAE-GAN:** El enfoque híbrido VAE-GAN busca combinar las fortalezas de los VAEs y las GANs para superar algunas de las limitaciones individuales de cada modelo. En este método:
 - Componentes Híbridos: Se utiliza la estructura del VAE para el proceso de codificación y decodificación, aprovechando la capacidad de este modelo para aprender una representación compacta y probabilística del espacio de datos. A su vez, se incorpora el mecanismo adversarial típico de las GANs.
 - Integración del Discriminador: En lugar de limitarse a la función de reconstrucción del VAE, se añade un discriminador que evalúa la calidad de las muestras generadas por el decoder (el “generador” en este contexto). Este discriminador ayuda a que el generador no solo se enfoque en minimizar la diferencia de reconstrucción, sino que también optimice la calidad perceptual y la adherencia a la distribución real de los datos.
 - Ventajas del Enfoque Híbrido: El enfoque híbrido ofrece ventajas significativas: la integración de un discriminador permite que los datos generados no solo retengan las propiedades estadísticas aprendidas mediante el VAE, sino que también imiten de forma más precisa la apariencia real de los datos de entrenamiento, mientras que la combinación de la función de reconstrucción del VAE con la pérdida adversarial de la GAN ayuda a mitigar problemas comunes como la inestabilidad y el colapso de modos que a menudo afectan a las GANs tradicionales.

4.6. MÉTODOS DE DESCOMPOSICIÓN Y ANÁLISIS ESPECTRAL DE SERIES DE TIEMPO

Para extraer componentes informativas que mejoren la capacidad predictiva de los modelos se emplearon tres técnicas de descomposición/transformación de la señal: STL, Wavelet y Multitaper. A continuación, se describen sus fundamentos teóricos:

- **Descomposición STL (Seasonal-Trend decomposition using LOESS):** Es un método para descomponer una serie de tiempo en sus componentes de tendencia, estacionalidad y residuo. STL utiliza un suavizado local ponderado secuencialmente (LOESS) para estimar estos componentes. Es importante destacar que STL utiliza una ventana de datos pasados para su cálculo.
- **Análisis Wavelet:** El análisis wavelet es una técnica que descompone una serie de tiempo en diferentes escalas de frecuencia. Esto se logra mediante la convolución de la serie de tiempo con una función base llamada "wavelet" a diferentes escalas y desplazamientos. Cada coeficiente wavelet resultante representa la correlación entre la serie de tiempo y el wavelet en una escala y posición específicas. Un aspecto crucial del cálculo wavelet es que el valor resultante se asocia con el centro del wavelet que se convoluciona a lo largo de la serie de tiempo.
- **Análisis Multitaper:** El análisis multitaper es una técnica espectral que se utiliza para estimar la densidad espectral de potencia de una serie de tiempo. En lugar de utilizar una única ventana de datos como en los métodos espectrales tradicionales, el multitaper utiliza un conjunto de ventanas ortogonales (tapers) para reducir la varianza de la estimación espectral y mejorar la detección de características espectrales. Al igual que STL, el análisis multitaper utiliza una ventana de datos pasados para su cálculo.

4.7. PREPROCESAMIENTO Y MANEJO DE DATOS

El preprocesamiento de datos es crucial en el análisis de series temporales financieras, facilitando la correcta interpretación y procesamiento por parte de los modelos predictivos. Los siguientes son los métodos de preprocesamiento usados en el proyecto:

- **Normalización y Estandarización:** La normalización, como la implementación de MinMaxScaler, asegura que las entradas de los modelos estén en una escala común, facilitando el aprendizaje eficiente por parte de los algoritmos. La estandarización ayuda a centrar los datos eliminando el promedio y escalando a la varianza unitaria, crucial para muchos modelos predictivos que suponen que los datos están centrados y normalizados [18].
- **Ingeniería de características:** El proceso de generación de nuevas características a partir de series de tiempo, como indicadores técnicos (RSI, MACD, Bollinger Bands), proporciona información adicional que puede ser crucial para la predicción. Estos indicadores ayudan a identificar tendencias y patrones en los datos que no son inmediatamente evidentes en los datos originales [19].

- **Extracción de características:** busca reducir la dimensionalidad y el ruido de la entrada para facilitar el entrenamiento de un modelo predictivo al identificar los posibles patrones en los datos originales y pasarlos ya procesados al modelo predictivo, además presentan la ventaja de que pueden crearse extractores de características adaptados al tipo de datos específico, pues los diferentes tipos de datos pueden contener patrones identificables por diferentes modelos de Deep Learning que se usan como extractores de características [21].

4.7.1. INDICADORES TÉCNICOS

- **RSI (Índice de Fuerza Relativa):** Mide la velocidad y el cambio de los movimientos de precios para evaluar condiciones de sobrecompra o sobreventa en un rango de 0 a 100.
- **MACD (Media Móvil de Convergencia/Divergencia):** Muestra la relación entre dos medias móviles de precios y se utiliza para identificar posibles cambios en la dirección de la tendencia.
- **MACD_Histogram:** Representa gráficamente la diferencia entre la línea MACD y su línea de señal, ofreciendo señales tempranas de posibles cruces.
- **MACD_Signal:** Es una media móvil de la línea MACD, utilizada para generar señales de compra o venta cuando cruza la línea MACD.
- **EMA (Media Móvil Exponencial):** Da mayor ponderación a los precios más recientes, haciéndola más sensible a los nuevos datos que una media móvil simple.
- **Stochastic_%K:** Indica la ubicación del precio de cierre actual en relación con el rango de precios durante un período determinado, útil para identificar posibles puntos de inflexión.
- **Stochastic_%D:** Es una media móvil de %K, suavizando sus fluctuaciones y proporcionando señales de trading más claras cuando cruza %K.
- **ADX (Índice Direccional Promedio):** Mide la fuerza de una tendencia, independientemente de su dirección, con valores más altos indicando tendencias más fuertes.
- **DI+ (Indicador Direccional Positivo):** Mide la fuerza de la tendencia alcista.
- **DI- (Indicador Direccional Negativo):** Mide la fuerza de la tendencia bajista.
- **ATR (Rango Verdadero Promedio):** Mide la volatilidad del mercado calculando el rango promedio de precios durante un período específico.
- **CCI (Índice de Canal de Materias Primas):** Mide la variación del precio de un activo con respecto a su promedio estadístico, identificando condiciones de sobrecompra o sobreventa.

4.8. MÉTRICAS DE PRECISIÓN

La precisión es el inverso del error en la predicción, también se debe medir la variación en las predicciones, estas medidas se pueden calcular de distintas maneras para facilitar la comparación con otros experimentos. Las medidas usadas en este proyecto son:

- **Error Cuadrático Medio (RMSE)** y Error Absoluto Medio (MAE): Estas métricas cuantifican el error de predicción, siendo críticas para evaluar la precisión de los modelos. Un MSE bajo y un MAE bajo indican un modelo que predice con alta precisión [2] [17].
- **Coefficiente de Determinación (R^2):** El R^2 mide la cantidad de varianza en las predicciones explicada por el modelo, proporcionando una medida de qué tan bien las predicciones se corresponden con los datos reales [2] [17].

4.9. MÉTRICAS DE ROBUSTEZ

La robustez es inversamente proporcional a estimaciones del **overfitting y underfitting**. El overfitting en modelos de deep learning ocurre cuando un modelo se ajusta demasiado a los datos de entrenamiento, capturando ruido y peculiaridades que no se generalizan bien a nuevos datos [2]. Mientras el **underfitting** ocurre cuando el modelo con posiblemente insuficientes datos de entrenamiento, no alcanza a ajustar sus parámetros adecuadamente [2]. Para estimar el overfitting/underfitting, se utilizan las siguientes métricas y técnicas:

- **Diferencia entre Error de Entrenamiento y Validación:** Comparar el error en los datos de entrenamiento con el error en los datos de validación es una técnica fundamental. Una gran disparidad entre estos errores sugiere overfitting, con el error de validación siendo significativamente mayor [17].
- **Variación del Coeficiente de Determinación (R^2):** En modelos de regresión, un R^2 alto en datos de entrenamiento, pero significativamente menor en datos de validación puede ser un indicio de overfitting, indicando que el modelo no generaliza bien [17].

4.10. CONCEPTOS DE TRADING RELEVANTES EN EL PROYECTO

- **Estrategia Heurística de Trading:** Método de toma de decisiones en el mercado financiero basado en reglas prácticas fundamentadas en el análisis de predicciones de precios, que busca identificar oportunidades de operación (compra o venta) evaluando la relación beneficio-riesgo bajo distintos horizontes temporales.
- **Backtrader:** Librería en Python especializada en simulación de estrategias de trading en condiciones que emulan el entorno de un broker real, permitiendo ejecutar operaciones con datos históricos.
- **Orden de Compra/Venta (Long/Short):** Ejecución de una operación financiera que implica adquirir (long) o vender (short) un activo, basada en la expectativa de que el precio subirá o bajará, respectivamente, optimizando el rendimiento esperado en función del análisis predictivo.
- **Riesgo (Risk):** Se utilizó el **Maximum Drawdown** como magnitud del riesgo y es la máxima pérdida asociada a una posición abierta antes de su cierre. En este proyecto se controla mediante límites de exposición (tamaño de orden), umbrales de stop loss y take profit, y el ratio beneficio/riesgo mínimo exigida antes de abrir una operación.
- **Tasa Ganancia/Riesgo:** Cociente entre el beneficio potencial estimado de una operación y la pérdida máxima proyectada. Este ratio se utiliza como criterio fundamental para determinar

la viabilidad de ejecutar una operación, asegurando que el retorno esperado justifique el riesgo asumido.

- Umbral de Rentabilidad Mínima: Valor mínimo aceptable para la tasa ganancia/riesgo, definido como parámetro configurable. Solo se consideran operaciones cuya rentabilidad esperada supere este umbral, con el objetivo de filtrar escenarios operativos desfavorables.
- Take Profit: Nivel predefinido de precio al cual se cierra automáticamente una posición con ganancias. Este valor se configura en función de los extremos previstos en la predicción de largo plazo y mediante factores ajustables que permiten adaptabilidad ante variaciones en la calidad predictiva.
- Stop Loss: Nivel predefinido de precio al cual se cierra automáticamente una posición con pérdidas controladas. Su ubicación se basa en los extremos negativos estimados en la predicción, ajustado mediante parámetros configurables que buscan proteger el capital ante movimientos adversos.
- Volumen de Orden: Cantidad de capital asignado a una operación específica. Se calcula considerando tanto el perfil de riesgo de la predicción como factores configurables, garantizando una exposición proporcional al nivel de confianza en la predicción.
- **Cierre Anticipado de Orden:** Mecanismo de gestión dinámica de posiciones que permite cerrar una operación antes de alcanzar los niveles de Take Profit o Stop Loss, cuando las condiciones del mercado —evaluadas mediante predicciones de corto y largo plazo— indican un deterioro en el escenario esperado.
- Reversión de Tendencia: Cambio anticipado en la dirección del mercado detectado mediante modelos predictivos. Su identificación permite cerrar operaciones con anticipación para evitar que una tendencia adversa comprometa la rentabilidad de la posición abierta.
- Optimización de Parámetros: Proceso mediante el cual se ajustan automáticamente los valores de configuración de la estrategia (como el umbral de rentabilidad, niveles de TP/SL, volumen, entre otros), utilizando algoritmos evolutivos —como el algoritmo genético implementado con la librería DEAP— con el fin de maximizar indicadores de rendimiento como el profit/risk ratio.
- Profit/Risk Ratio: Indicador de eficiencia operativa definido como la razón entre el balance final obtenido y el balance mínimo alcanzado durante la simulación, partiendo de un capital inicial determinado. Se emplea como métrica de referencia para evaluar la robustez y sostenibilidad de la estrategia.

5. ANTECEDENTES

Los trabajos revisados en esta sección proporcionan una base sobre la utilización de modelos de Deep-Learning y la incorporación de datos sintéticos en la predicción de series temporales financieras. El proyecto propuesto se distingue por predicción de corto, mediano y largo plazo, y por evaluar el impacto de diferentes cantidades y tipos de datos de entrenamiento, incluyendo datos sintéticos, en la precisión y robustez de los modelos predictivos. Estas diferencias enfatizan la originalidad y relevancia del proyecto actual.

El Artículo "Financial Time Series Models—Comprehensive Review of Deep Learning Approaches and Practical Recommendations"[10] realiza una revisión exhaustiva de los enfoques de deep learning aplicados a la predicción de series temporales financieras. Analiza 15 artículos destacados publicados en los últimos siete años, comparando sus modelos y métodos de predicción. Los modelos se entrenaron para predecir, con un día de antelación, el valor de 29 índices y precios de acciones y commodities en cinco periodos diferentes. El estudio destaca la falta de consenso sobre el mejor enfoque y ofrece recomendaciones prácticas sobre cuándo utilizar cada modelo en función de los datos disponibles.

Los datos utilizados en los estudios revisados incluyen una amplia gama de series temporales financieras, como precios de acciones, índices y datos de commodities. Los autores discuten diversas técnicas de preprocesamiento de datos, incluyendo la normalización, la eliminación de valores atípicos y el uso de técnicas de suavizado para mejorar la calidad de los datos de entrada. Además, se destacan enfoques para integrar datos heterogéneos, como indicadores técnicos y análisis de sentimiento, para enriquecer los conjuntos de datos y mejorar la capacidad predictiva de los modelos.

Este trabajo aporta una revisión crítica y comparativa de múltiples usos de deep learning en la predicción financiera, señalando las dificultades en la replicación de los resultados debido a la falta de datos y códigos accesibles. Aunque proporciona recomendaciones útiles, no aborda específicamente la segmentación de predicciones en diferentes horizontes temporales ni el uso de datos sintéticos. Este aspecto es fundamental en el proyecto actual, que evaluó el impacto de diferentes cantidades y tipos de datos en la precisión y robustez de los modelos predictivos. [10]

En el artículo "Generation of Realistic Synthetic Financial Timeseries" [14] se investigan varias soluciones para la generación de series temporales financieras sintéticas utilizando modelos generativos profundos. Los autores proponen el uso de GANs, autoencoders variacionales y redes de coincidencia de momentos generativos para capturar las propiedades específicas de las series temporales financieras, como la distribución de colas gruesas, la autocorrelación específica y la volatilidad en clústeres. Se llevaron a cabo experimentos con datos reales del mercado de valores de EE.UU. para validar el rendimiento de los modelos generativos propuestos.

Los datos utilizados incluyen precios históricos de acciones y otros activos financieros del mercado de valores de EE.UU. El preprocesamiento de datos implicó la normalización y la segmentación en ventanas temporales para facilitar el entrenamiento de los modelos generativos. Los autores también utilizaron técnicas de enriquecimiento de datos mediante la integración de indicadores

técnicos y métricas de volatilidad para crear conjuntos de datos sintéticos que reflejen las propiedades estadísticas de los datos reales.

Los modelos generativos fueron evaluados utilizando métricas cualitativas y cuantitativas, incluyendo la divergencia de Kullback-Leibler y la métrica de Frechet para evaluar la calidad de las series sintéticas en comparación con los datos reales. Los resultados mostraron que los modelos generativos podían replicar efectivamente las propiedades estadísticas de las series temporales financieras, lo que es crucial para mejorar la robustez de los modelos predictivos cuando se utilizan datos sintéticos.

El principal aporte de este trabajo es la demostración de que los modelos generativos profundos pueden crear series temporales financieras sintéticas que capturan eficazmente las propiedades estadísticas de los datos reales. Este estudio también introduce métricas cualitativas y cuantitativas para evaluar la calidad de las series sintéticas, lo cual es relevante para el proyecto actual. Sin embargo, no se enfoca en la segmentación de las predicciones en corto y largo plazo, ni en la evaluación del impacto de diferentes cantidades de datos. [14].

El artículo "Financial Time Series Forecasting with the Deep Learning Ensemble Models" [24] propone un modelo de ensamblaje que combina CNN, LSTM y ARIMA para la predicción de series temporales financieras. Utiliza un enfoque híbrido donde cada modelo contribuye con sus fortalezas: el CNN-LSTM para capturar características espaciotemporales y el ARIMA para modelar la autocorrelación de los datos. Los datos utilizados incluyen precios históricos de acciones, índices y otros indicadores financieros.

Se implementó un riguroso proceso de preprocesamiento de datos que incluyó la normalización y la eliminación de valores atípicos para mejorar la calidad de los datos de entrada. Además, los autores aplicaron técnicas de enriquecimiento de datos mediante la inclusión de indicadores técnicos como medias móviles y el índice de fuerza relativa (RSI) para proporcionar una mayor riqueza de información al modelo. El modelo se evaluó utilizando la métrica de error cuadrático medio (RMSE) y mostró una reducción significativa del error en comparación con otros modelos convencionales, logrando así una mayor precisión y robustez en las predicciones.

El aporte principal de este trabajo radica en la integración de múltiples arquitecturas de Deep-Learning para aprovechar sus capacidades complementarias y sus ventajas específicas para diferentes tipos de datos. Sin embargo, este estudio no segmenta sus predicciones en corto y largo plazo, ni evalúa específicamente el impacto de diferentes cantidades de datos en cada horizonte temporal, como se propone en el proyecto actual. Además, aunque se incluyen indicadores técnicos, no se profundiza en el uso de datos heterogéneos como indicadores fundamentales.

6. METODOLOGÍA

Para llevar a cabo la investigación sobre la predicción de la tasa de cambio EUR/USD mediante modelos de Deep Learning, se usó una metodología estructurada que comprende la recopilación, preprocesamiento, modelado y evaluación de datos. A continuación, se detallan los principales componentes del diseño del desarrollo.

6.1. DATASETS DESCARGADOS

Se recopilaron los datos de tasa de cambio de EUR/USD con una periodicidad de una hora desde Kaggle. Los dataset que se descargaron se refieren a continuación:

- “EURUSD Forex Pair Historical Data (2002 - 2020)”. Según la documentación proviene del bróker Oanda, pero contiene datos desde 2005, no 2002 para las columnas Open, Close, High y Low. A continuación, se hace una definición corta de lo que representa cada columna de la base de datos, como se mencionó, el intervalo temporal es de una hora.
 - Date: Fecha y hora del registro. Campo de formato fecha
 - Open: Precio de apertura del par EUR/USD para cada intervalo temporal. Representa el primer precio registrado al inicio del periodo correspondiente.
 - Close: Precio de cierre del par EUR/USD para cada intervalo temporal. Indica el último precio registrado al finalizar dicho periodo.
 - High: Precio máximo alcanzado por el par EUR/USD durante el intervalo analizado. Refleja el punto más alto de volatilidad positiva en ese periodo.
 - Low: Precio mínimo registrado por el par EUR/USD en el intervalo considerado. Marca el nivel más bajo alcanzado por la cotización en ese lapso.

La data cuenta con 93.084 filas. Dado que la información pertenece a señales de mercado de valores, en este caso por negocio no se tiene información para los fines de semana. En este dataset se aplicaron algunos análisis de estacionalidad y detección de señales que se verificarán más adelante.

- Para indicadores fundamentales (S&P 500 y Volatility Index) se usó el dataset llamado: “S&P 500 Historical Data” y “VIX (CBOE Volatility Index) Daily Updated” que contienen datos con periodicidad diaria (cada 24 horas un dato) provenientes de Yahoo Finance. Estos conjuntos de datos contienen 3 974 registros diarios cada uno entre el 03-ene-2005 y el 31-dic-2020. Las columnas disponibles son Date, Open, Close, High y Low. Para este estudio cada valor diario se replicó a las horas de negociación pertinentes, generando 55 318 observaciones horarias alineadas con el set EUR/USD.
- Para la prueba de alta frecuencia, se usó el dataset: “Forex Top Currency Pairs (2000-2020)”, que contiene información de particionada cada 15 minutos. Este dataset tiene las mismas columnas que la tabla principal (Date, Open, Close, High y Low) y se refiere a los datos de las tasas de cambio Forex de los pares euro versus dólar.

- Como dataset auxiliar para la selección de periodicidad se utilizó “Market Stocks - M1 Ticks” desde Kaggle que contiene los datos provenientes de la plataforma MetaTrader con granularidad de 1 minuto, de este se utilizó la columna Close únicamente solo para construir series agregadas a 5 min, 15 min, 1h, 4h y 24h mediante down-sampling y con ellas calcular las métricas de Teoría de la información descritas en la sección 5.3.

Nota: Las transformaciones, la creación de características y la generación de los conjuntos de entrenamiento, validación y prueba se describen en las secciones 8 y 9 del documento.

6.2. ARQUITECTURA GENERAL DEL SISTEMA

El sistema está compuesto por módulos independientes y configurables, diseñados para manejar datos de entrada de diversas fuentes, preprocesarlos y entrenar modelos de predicción. Estos módulos fueron implementados en Python y estructurados de la siguiente manera:

- **Módulo de Preprocesamiento de Datos:** Se encarga de la limpieza, normalización y selección de características. Utilizó la biblioteca scikit-learn para aplicar normalización min-max y guardar los parámetros como un archivo JSON con los parámetros de normalización que se usan para desnormalizar las predicciones.
- **Extractor de Características:** Fue implementado con un autoencoder para la reducción de dimensionalidad produciendo características que contienen patrones relevantes (características extraídas).
- **Módulo de Predicción:** Es configurable usando el sistema built-in de plugins de Python (setup.py), con tres tipos de modelos de Deep Learning: Convolutional Neural Networks (CNN), Long Short-Term Memory Networks (LSTM) y Transformers, desarrollado con el framework Keras, adicionalmente se realizaron plugins para ANN, para el modelo base Random Forest y para todos los modelos híbridos usados.
- **Módulo de Ingeniería de Características:** Genera indicadores técnicos, datos estacionales y convierte datos fundamentales en representaciones listas para ser adicionadas como canales adicionales a la serie de tiempo base de periodicidad horaria del EURUSD que usan los predictores, en el rango de fechas común, utilizando pandas.
- **Generador de Datos Sintéticos:** Se implementó para aumentar la cantidad de datos de entrenamiento mediante técnicas de generación de datos artificiales, con plugins para VAE, y Conditional VAE.
- **Estrategia Heurística de Trading:** Para determinar la robustez de las predicciones realizadas en un entorno realista, se programó una estrategia ideal de trading, simulada en Backtrader, que permite conocer las condiciones para tener una ganancia o pérdida establecida. Esto es útil porque permite tener una comparativa del rendimiento del modelo adicional a los indicadores convencionales. La herramienta toma las predicciones de corto y largo plazo para producir una secuencia de ordenes de mercado que determinan a partir de un balance inicial de USD 10.000 tanto el balance final como el mínimo balance para producir una métrica de profit/risk de las predicciones usadas incluyendo comisiones por apertura y cierre de operaciones.
- **Sistema de Configuración Integrado:** Todos los módulos permiten guardar (al usar parámetros via CLI) y cargar archivos JSON con la configuración específica usada para todos

los hiperparámetros y parámetros relevantes del entrenamiento de cada modelo, lo que garantiza la replicabilidad de los experimentos. Esto también permitió la ejecución secuencial automática de múltiples experimentos pre-configurados por medio de scripts en Windows (y Linux). Tanto configuraciones JSON, como datos, resultados y scripts de ejecución por lotes, se incluyen en la carpeta examples de todos los repositorios de código.

6.3. SELECCIÓN DE PERIODICIDAD PARA PREDICCIONES

En esta sección se describe como se seleccionaron las periodicidades de corto y largo plazo de los datasets a usar para entrenar y realizar inferencia en los modelos.

Para evaluar que datos proveen menos ruido respecto a la señal subyacente de precio, se calculó una serie de métricas estadísticas y otras basadas en Teoría de la Información que permiten medir tanto el ruido presente como la información libre de ruido de la señal, para ayudar a seleccionar las periodicidades de corto y largo plazo a usar en las predicciones. Las métricas definidas en la Sección 3.3 se aplicaron a las frecuencias de 5 min, 15 min, 1 h, 4 h y 24 h, obteniendo los resultados mostrados en la Tabla 1.

Para este análisis se descargó un dataset con periodicidad de un minuto desde Kaggle llamado: "Market Stocks - M1 Ticks" que contiene datos de la plataforma MetaTrader de 2003 a 2024 y se realizó downsampling a las periodicidades de: 5 minutos, 15 minutos, una hora, 4 horas, y 24 horas, las periodicidades se convirtieron a segundos para usar el **Teorema de Capacidad de Canal de Shannon/Hartley**, que requiere el SNR(S/N), la periodicidad (1 / frecuencia B) y la cantidad de datos en la serie de tiempo convertida a el tiempo en segundos T, para calcular la información libre de ruido del dataset con la fórmula:

$$I = C \times T = T \times B \log_2 \left(1 + \frac{S}{N} \right)$$

donde:

- I es la información total en bits,
- T es el tiempo en segundos,
- B es el ancho de banda en hertz,
- $\frac{S}{N}$ es la relación señal a ruido.

Estos cálculos se realizaron utilizando una hoja de datos que consolida las estadísticas e indicadores de cada conjunto. La Tabla 1, fue clave para identificar los conjuntos de datos a usar en el resto del proyecto.

channel_capacity[bits/s]	information_bits	NR=ruido/retornos	periodicidad_segundos	Information/RR
0,3121	84278,5121	14,1364	60	5961,7938
0,0422	56982,7200	46,7021	300	1220,1318
0,0149	60333,1208	23,7403	900	2541,3773
0,0036	58393,0454	16,0151	3600	3646,1257
0,0007	43117,0669	22,8225	14400	1889,2371
0,0001	57398,0603	4,0027	86400	14339,9241

Tabla 1 - Estadísticas para Selección de Periodicidades de Corto y Largo Plazo

En la Tabla 1, se realizaron las siguientes observaciones:

- El dataset de 24 horas (86400 segundos) presentó el menor valor de RR (4.0027), lo que indica un nivel mínimo de ruido en relación con los retornos (variaciones por tick), y también posee el valor más alto de NR (14339.9241), destacándose por su estabilidad y alta cantidad de información útil para modelos que priorizan tendencias de largo plazo.
- El dataset de 1 hora (3600 segundos) mostró un balance entre bajo ruido y estabilidad, con un RR moderado (16.0151) y un NR de 3646.1257, haciéndolo un buen candidato para capturar dinámicas más detalladas de la tasa de cambio sin comprometer excesivamente la precisión debido al ruido. Aunque los datos de 60 segundos tienen un mayor valor de RR que la predicción de una hora, por limitaciones de procesamiento y hardware, se opta por el dataset de una hora.

Una vez determinada la frecuencia óptima, los experimentos de modelado se realizaron con el dataset EUR/USD de 1 hora descrito en la **Sección 5.1**, que sí contiene las columnas Open, High, Low y Close necesarias para las fases posteriores.⁹

Debido a estos hallazgos se selecciona la periodicidad de **una hora para las predicciones de corto plazo y de 24 horas para el largo plazo**. Como se mostrará en las siguientes secciones, estas periodicidades no son el horizonte predictivo, sino la periodicidad de las predicciones. Como se verá más adelante, se usaron 6 horas y 6 días (144 horas) como horizontes de predicción de corto y largo plazo respectivamente, el razonamiento para la selección del horizonte predictivo se muestra en la sección 6.4 contiene datos con periodicidad de 15 minutos obtenidos desde la plataforma MetaTrader.

6.4. HORIZONTE PREDICTIVO

La determinación del horizonte predictivo de corto y largo plazo es una decisión crítica en el diseño de estrategias de trading. A continuación, se presentan las razones prácticas, normativas y estratégicas que justifican la elección de horizontes predictivos de corto y largo plazo en función de una estrategia de trading que usa reglas empíricas basadas en predicciones ideales. Esta estrategia heurística se describe en la sección 6.7 y fue implementada usando la librería Backtrader, se verificó su correcto funcionamiento usando predicciones ideales antes de usarse con las predicciones reales producidas por los predictores de este proyecto.

6.5. DETERMINACIÓN DE HORIZONTE TEMPORAL DE LARGO PLAZO

En el largo plazo, el objetivo es identificar en las predicciones, el mínimo, el máximo y también el orden en que aparecen estos valores, de manera que con la predicción de largo plazo se determine el potencial beneficio/riesgo de una orden de venta o compra, esto determina tanto los momentos óptimos para abrir órdenes, como el tipo de orden a abrir (buy/sell) y también su configuración de Take Profit y Stop Loss.

Estas predicciones se usan cada vez que se cierra una orden para buscar tendencias estables y no se usará un plazo más largo de 10 días hábiles (2 semanas) para evitar riesgos asociados a factores fundamentales que puedan afectar significativamente la tendencia. Además se debe tener en cuenta la legislación de Estados Unidos en el límite de transacciones en cuentas de margen, como las que se usan en los brokers de Forex, esta legislación comúnmente llamada límite de **Pattern Day Trading (PDT)** es realizada por el Financial Industry Regulatory Authority (FINRA) y dice que se pueden realizar hasta **3 trades en una ventana de 5 días**, de lo contrario, el cliente debe tener USD 25.000 de equity (balance + profit de ordenes abiertas) o las ordenes no serán ejecutadas.

En otras palabras, usando el **PDT**, se determinó que no se debe abrir más de una orden pasando un día, por tanto con una estrategia conservadora, se busca que se realicen entre 1 y 2 órdenes por semana (**baja frecuencia**), y por tanto se considera que un buen horizonte predictivo para esta meta es 6 días, que cubren una semana y el primer día de la siguiente, tiempo en el que la estrategia propuesta busca en las predicciones el mejor trade a realizar en condiciones ideales, si se tuviesen predicciones perfectas. También se realizaron simulaciones de **alta frecuencia** para contrastar.

6.6. DETERMINACIÓN DE HORIZONTE TEMPORAL DE CORTO PLAZO

Para el corto plazo, el enfoque está en verificar con la mayor frecuencia posible, si cerrar órdenes abiertas antes de alcanzar el Stop Loss, que es un umbral predefinido de pérdida que cierra automáticamente una operación para limitar el riesgo. En la estrategia heurística ideal propuesta, las predicciones de los modelos de Deep learning se actualizan cada hora con un horizonte de 6 a 12 horas. Las predicciones de este rango de tiempo permiten reaccionar rápidamente a los cambios del mercado y eventos de calendario económico o factores geopolíticos que puedan producir cambios abruptos en la tendencia y volatilidad, optimizando la gestión del riesgo y asegurando flexibilidad en las decisiones, por lo tanto, para este enfoque utilizaremos un horizonte predictivo intermedio de 6 horas.

Además, la predicción a corto plazo permite detectar si no se logrará el Take Profit (objetivo de ganancia previamente establecido) de una orden abierta en el plazo estimado cuando se abrió, de manera que se pueda cerrar la orden si se detecta que se ha llegado a un valor cercano al Take Profit y se predice una reversión de tendencia, que evitará alcanzar este valor.

6.7. ESTRATEGIA HEURÍSTICA DE TRADING BASADA EN PREDICCIONES DE CORTO Y LARGO PLAZO

Esta estrategia heurística fue implementada usando la librería **Backtrader** en Python que permite simular en condiciones similares a las de un Broker, con parámetros configurables para diversos factores que permiten simular condiciones de la vida real. Si bien la herramienta hace posible calcular escenarios, no permite hacer predicciones. Las simulaciones permiten tener los escenarios donde se hubiese podido invertir de la mejor manera según el comportamiento de la serie en ese momento del tiempo. Esta estrategia está diseñada para manejar una secuencia de órdenes, una sola en cada momento y no maneja múltiples ordenes paralelas. A continuación, se describen las tácticas de apertura y cierre de órdenes que componen la estrategia, estas son configurables por medio de los parámetros descritos, que fueron optimizados con un algoritmo genético usando la

librería **DEAP** en Python para maximizar el profit/risk ratio (balance final / mínimo balance empezando con USD 10.000) en la mejor predicción de corto y largo plazo obtenida entre todos los predictores usados, lo que permite usarla como base para comparar su desempeño en trading respecto a las demás predicciones.

6.7.1. APERTURA DE ÓRDENES DE COMPRA O VENTA (SHORT/LONG)

La estrategia se fundamenta en el análisis de predicciones a corto y largo plazo, evaluando en primer lugar los máximos y mínimos proyectados, así como los valores extremos previos a estos valores, para determinar el potencial de ganancia y el riesgo asociado a una posible operación. Con esta información se calcula la tasa ganancia/riesgo para identificar la viabilidad tanto de una orden de compra como de venta.

Posteriormente, se compara la tasa ganancia/riesgo de una orden de compra con la de una de venta y se selecciona aquella con el mejor balance entre beneficio potencial y riesgo, siempre y cuando la tasa ganancia/riesgo supere un umbral predefinido configurable.

La predicción a largo plazo también se usa para la configuración de la orden a realizar, determinando el Take Profit, el Stop Loss y el Volumen de la orden basándose en los valores extremos encontrados en la predicción y en factores configurables para estos parámetros que proporcionan flexibilidad ante predicciones no-ideales. De esta forma, se garantiza que sólo se ejecuten operaciones con una relación favorable entre la ganancia esperada y la posible pérdida.

6.7.2. CIERRE DE ORDENES

Una vez abierta una posición, la estrategia utiliza las **predicciones de corto plazo** para monitorear el comportamiento inmediato del mercado, está programada para realizar este monitoreo con frecuencia de una hora. Si se detecta que, en el corto plazo, el precio se aproxima al nivel de **Stop Loss**, se procede a cerrar la operación de forma anticipada, protegiendo así el capital invertido ante un posible deterioro de la situación por factores que alteran la tendencia como eventos económicos o geopolíticos de gran impacto.

Asimismo, las predicciones tanto de corto como de largo plazo se utilizan para evaluar si el precio alcanzará el nivel de Take Profit dentro del horizonte temporal previsto, sin que se produzca un cambio inesperado en la tendencia del mercado. Si alguna de las predicciones indica que el objetivo de ganancia no se concretará antes de un posible cambio de dirección, la operación se cierra anticipadamente con el fin de evitar pérdidas innecesarias

Las condiciones de cierre, también se hacen usando factores configurables (optimizables) para los límites a alcanzar antes de un cierre anticipado.

6.8. EXPERIMENTOS PARA CUMPLIR OBJETIVOS

En las siguientes secciones se muestran los pasos realizados para realizar los experimentos requeridos para cumplir cada uno de los objetivos específicos propuestos. En la sección 7 se muestra la exploración preliminar de los datos, en la sección 8 se muestra su preparación y

preprocesamiento incluyendo la ingeniería de características, en la sección 9 se describen los modelos usados en todos los experimentos, en la sección 10 describe el procedimiento y resultados de variar el número de registros de entrenamiento, en la sección 11 se muestra el procedimiento y resultados de variar el número y tipo de características usadas durante el entrenamiento, en la sección 12 se evalúa el impacto del uso de una reducción de características usando autoencoder y en la sección 13 se usan datos sintéticos producidos con diferente técnicas para verificar si contribuyen a la mejora en precisión de los predictores.

Bloque experimental	Dataset / tamaño (entrenamiento)	Variación principal	Modelos evaluados	Sección
Cantidad de datos (<i>línea base</i>)	EUR/USD 1 h – 25 200 filas	8 características básicas (OHLC + derivadas)	ANN · CNN · LSTM · Transf.	11
Cantidad de datos (<i>máximo histórico</i>)	EUR/USD 1 h – 55 318 filas	8 básicas	ANN · CNN · LSTM · Transf.	11
Variación de características	1 h – 25 200 filas	+ 11 componentes de decomposición (STL, Wavelet, Multitaper)	ANN · CNN · LSTM · Transf.	12.1
"	"	+ 12 indicadores técnicos (MACD, RSI, ...)	"	12.1.1
"	"	+ 3 variables estacionales (hora, día, mes)	"	12.1.2
"	"	+ 16 valores de alta frecuencia (últimos 8 ticks 15 m y 30 m)	"	12.1.3
"	"	+ 2 fundamentales (S&P 500, VIX)	"	12.1.4
"	"	Todas las anteriores (44 features)	"	12.1.5
Extractores sin preentrenamiento	1 h – 25 200 filas	Módulo extractor (ANN/CNN/LSTM/Transf.) + predictor fijo	ver nota ¹	12.2
Extractores pre-entrenados (Autoencoder)	1 h – 25 200 filas	Extractor ANN/CNN/LSTM/Transf pre - trained con autoencoder	ver nota ¹	12.3
Datos sintéticos	1 h base + hasta 8 años generados (VAE)	Volúmenes 2, 4 y 8 años extra a los 4 años reales	CNN · LSTM · Transf.	12.4

Tabla 2 - Consolidado de experimentos para cumplir objetivos

Notas:

¹ En el Capítulo 13 el “predictor fijo” es común a todas las pruebas con capas: 2xConv1D → LSTM → Bayesiana. Solo cambia la arquitectura del extractor que le precede y si está o no pre-entrenado con un autoencoder.

- Los horizontes de predicción son los mismos en todos los bloques: **6 h** (corto plazo) y **6 días** (largo plazo). Estas predicciones son usadas en las simulaciones de trading.
- Las métricas comparadas (MAE, RMSE, MAPE, SNR) y la simulación de trading aparecen detalladas en cada sección citada.

7. EXPLORACIÓN DE LOS DATOS

El análisis exploratorio permitió comprender mejor las características de los datos, se calcularon estadísticas como media, mediana, desviación estándar y SNR para cada característica de cada horizonte elegido, para verificar su integridad.

media	desviacion	snr	promedio_retornos	periodicidad_segundos
1,1233	0,0017	434366,9869	0,0001	60
1,1006	0,0137	6483,7115	0,0002	300
1,0566	0,0101	10863,6958	0,0003	900
1,1025	0,0123	8057,1464	0,0006	3600
1,1483	0,0415	765,1964	0,0013	14400
1,1221	0,0135	6912,1356	0,0024	86400

Tabla 3 - Análisis exploratorio básico de los datos

En la tabla 3 se muestran las métricas para las diferentes periodicidades de los conjuntos de datos utilizados, en particular para la periodicidad horaria se puede observar que la desviación estándar es aproximadamente un 1.2% de la media, indicando que los valores varían aproximadamente en este rango y coincidiendo con el promedio de retornos que se encuentra en 0.06%, un valor mucho menor a la desviación estándar, lo que indica una presencia de ruido baja como se puede confirmar con un valor alto en el SNR.

La Figura 1 muestra la distribución de la variable objetivo de predicción, conocida por ser el cierre de operación o variable “close”. En el contexto del análisis de series de tiempo financieras, el precio de cierre representa el último precio al que se negoció un activo, en este caso el tipo de cambio EURO/USD. Para tener una mejor identificación del comportamiento de la serie en el periodo desde 2005-02 hasta 2020-02, la figura se realiza con un formato diario.

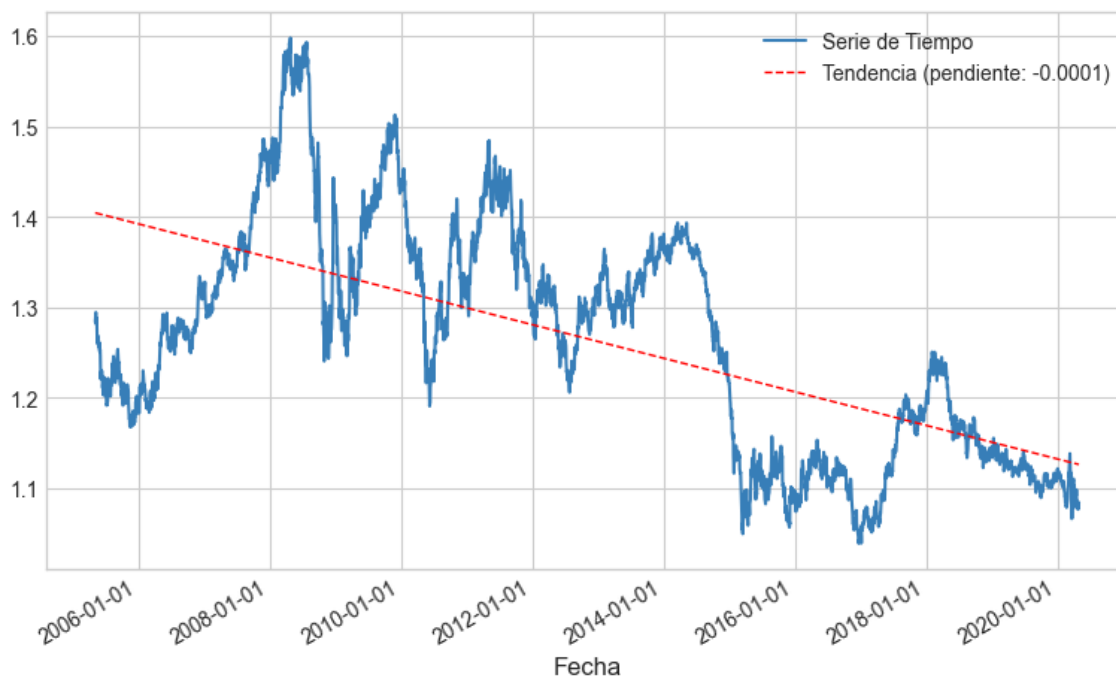


Figura 1 - Serie de tiempo con valores de cierre a 24 horas

Los datos trabajados revisados cada 24 horas se obtienen un total de 3881 observaciones. El valor promedio de cierre es de 1.266, con una desviación estándar de alrededor de 0.127, lo que indica una dispersión moderada de los datos alrededor de la media. El valor de cierre más bajo registrado es 1.039 y el más alto es 1.598. La mitad de los valores de cierre se encuentran por debajo de 1.274, mientras que el 25% más bajo es inferior a 1.140 y el 25% más alto supera 1.357. En general, la variable de cierre presenta una distribución relativamente centrada alrededor de 1.27, con una extensión que va desde aproximadamente 1.04 hasta 1.60. Sin embargo, como se observa en la Figura 1, los datos tienen una tendencia decreciente a lo largo del tiempo aplicando una regresión lineal simple a los mismos.



Figura 2 - Serie de tiempo con valores de cierre a una hora

Al hacer análisis de los datos de una hora hay un total de 93,084 observaciones. Su valor promedio es de aproximadamente 1.266, con una desviación estándar de alrededor de 0.127, lo que también sugiere una dispersión moderada de los datos en torno a la media. El valor mínimo registrado es 1.036 y el máximo alcanza 1.601. El 50% de los valores de la serie se sitúan por debajo de 1.274, mientras que el cuartil inferior se encuentra en 1.140 y el cuartil superior en 1.358. En términos generales, la serie tiene una distribución centrada alrededor de 1.27, con un rango que se extiende desde aproximadamente 1.04 hasta 1.60. En la Figura 2, los datos tienen una tendencia decreciente a lo largo del tiempo aplicando una regresión lineal simple, lo cual se espera dado que los datos diarios vienen de un filtro de los datos de hora.

Al realizar el muestreo de la información con periodicidad de 24 horas, realizando el filtro exactamente en el cambio de día y asumiendo que ese es el dato representativo, no se obtienen diferencias significativas en cuanto a media y mediana, quedando virtualmente en el mismo número y con variaciones en los valores después de la cuarta cifra. Los datos de desviación estándar

y rangos intercuartílicos son también muy similares, aunque los tamaños de muestra son notablemente menores para los datos con periodicidad de 24 horas (3.881) con respecto a los de hora (93.084)

Se realizó una descomposición usando **Seasonal and Trend decomposition using Loess (STL)** en la sección 7.1 para visualizar el comportamiento de los datos en el tiempo y algunas gráficas que pueden llegar a ser de utilidad para la visualización del comportamiento y entendimiento de sus cambios a través del histórico.

7.1. DESCOMPOSICIÓN CON STL

La descomposición STL es una técnica para separar una serie de tiempo en sus componentes fundamentales: tendencia (la dirección a largo plazo), estacionalidad (patrones cíclicos repetitivos dentro de un período fijo) y residuo (el ruido o las fluctuaciones irregulares restantes). Aislar los componentes puede brindar una visión más clara de la estructura de la serie temporal y puede brindar información relevante de forma visual. La Figura 3 muestra la descomposición de la serie en formato diario, donde en la primera gráfica se encuentra la gráfica de la serie de tiempo original, en un segundo espacio se encuentra la información de la tendencia, después el componente estacional y por último los residuos.

Tendencia, Estacionalidad y Residuales Largo Plazo

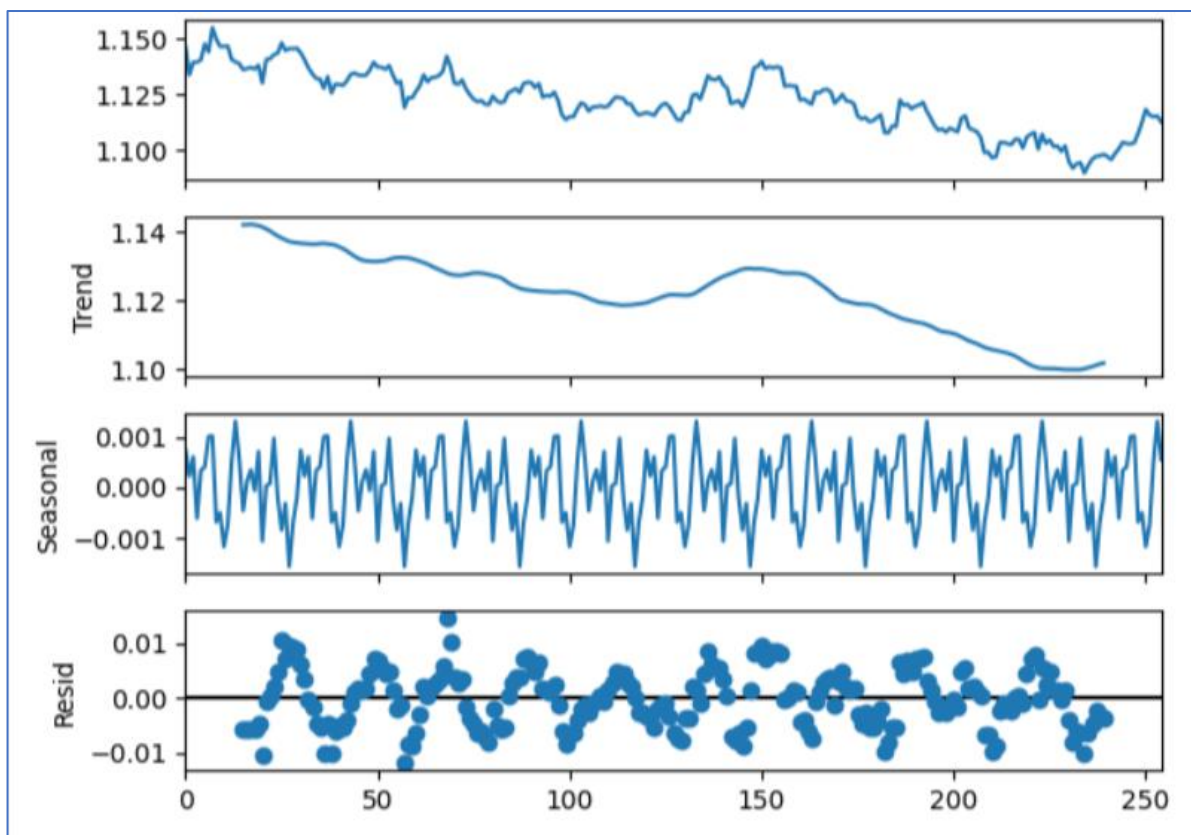


Figura 3 - Descomposición STL de periodicidad largo plazo

En la Figura 3 se puede visualizar la información de la tendencia, la estacionalidad y los residuales para la serie de tiempo de largo plazo. La tendencia muestra el patrón a largo plazo en los datos, para este caso particular se puede ver una tendencia decreciente a largo plazo lo que concuerda con el dato que arroja la línea de regresión vista en la Figura 1. Se observa también el gráfico de estacionalidad que refleja los patrones repetitivos que ocurren en intervalos regulares dentro de la serie de tiempo, con picos y valles lo cual puede representar una relación entre días de la semana. En el gráfico de la Figura 4 se puede observar el cálculo de este comportamiento para datos de 24 horas.

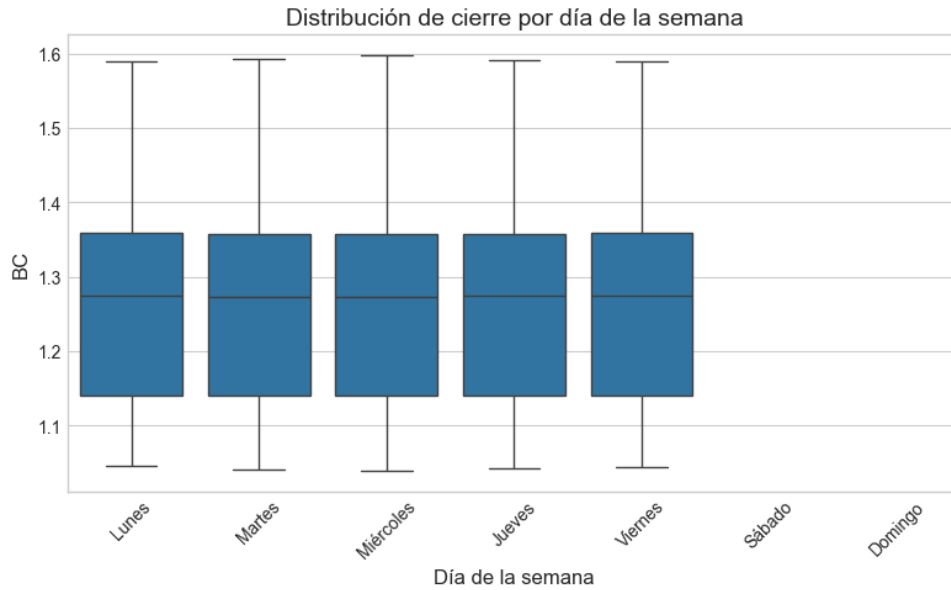


Figura 4 - Distribución de la serie de cierre por día de la semana

Aunque el comportamiento de la serie en las gráficas STL puede intentar evidenciar un comportamiento estacional, en la gráfica de boxplot de la figura no se puede evidenciar un cambio importante que sea consistente en el tiempo y en el cual los días de la semana sean factores que impacten en el análisis. En la Figura 5 se puede evidenciar el comportamiento de la distribución de los datos en cada mes del año. No se evidencia un comportamiento marcado, aunque sí se logra ver una mayor mediana de venta en los meses de inicio del año comparado con los meses finales, por ejemplo, en abril hay una mediana de venta mayor en comparación con octubre o noviembre.

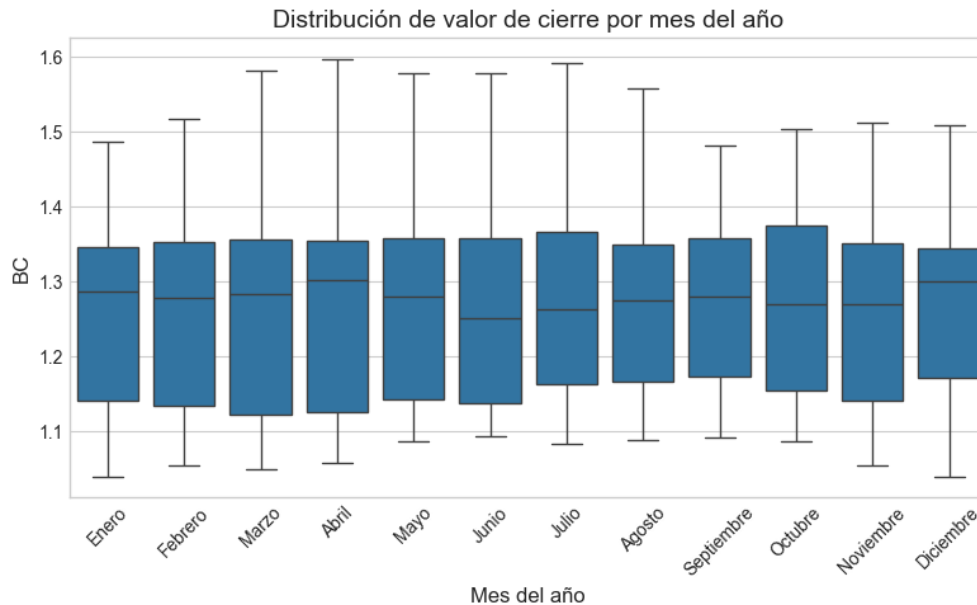


Figura 5 - Distribución de la serie de cierre por mes del año en formato diario

8. INGENIERÍA DE CARACTERÍSTICAS

Se procesaron los datasets descargados para generar los que se usarán con los predictores para los experimentos de variación en número y tipo de características de entrenamiento, se generaron en un rango de fechas común con el de EURUSD de periodicidad horaria. Se generaron datasets para indicadores técnicos, indicadores fundamentales y valores de alta frecuencia por tick.

- Se calcularon 12 **indicadores técnicos** frecuentemente usados como RSI, MACD, MACD_Histogram, MACD_Signal, EMA, Stochastic_%K, Stochastic_%D, ADX, DI+, DI-, ATR, CCI. A estos indicadores técnicos se les calculó el Coeficiente de Variación, que es el inverso cuadrado de la relación señal a ruido (SNR) y con el objetivo de usar señales bajas en ruido, se estableció un umbral de 1 para el CV ya que en este caso la señal domina y la serie presenta baja dispersión relativa, por lo tanto se dejaron solo las señales correspondientes a indicadores por debajo de este umbral ($SNR > 1$).
- Se procesaron los datasets de datos fundamentales que tienen periodicidad diaria para convertirlos a periodicidad horaria, usando el mismo valor para todas las horas del día. Se trabajaron un total de 3 indicadores fundamentales entre ellos WilliamsR, Momentum y ROC, de los cuales se hace una breve descripción a continuación:
 - WilliamsR (Williams %R): Mide el nivel del precio de cierre actual en relación con el rango máximo-mínimo durante un período determinado, oscilando entre 0 y -100 para identificar condiciones de sobrecompra o sobreventa.
 - Momentum: Mide la tasa de cambio del precio de un activo durante un período específico, indicando la velocidad con la que el precio está subiendo o bajando.
 - ROC (Rate of Change): Calcula el porcentaje de cambio entre el precio actual y el precio de hace un cierto número de períodos, mostrando la fuerza y la dirección del momentum.
- Se generaron **datasets de alta frecuencia por cada hora**, usando los últimos 8 ticks de 15 minutos y los últimos 8 ticks de 30 minutos del CLOSE de EURUSD para cada hora. Para poder generar esta información, se hace un shift de 8 al dataframe que contiene datos de 15 minutos. Esto hace que se tenga información en una misma línea y que puede llegar a nutrir los modelos de una mejor manera. En este caso específico, un desplazamiento de 8 períodos incorpora datos de 2 horas previas (8 períodos x 15 minutos/período) para cada instante, lo que permite que el modelo considere patrones temporales recientes al realizar sus predicciones. Se realiza el mismo procedimiento de desplazamiento de 8 períodos para los datos de 30 minutos.
- **Variables de estacionalidad:** Se añadieron columnas que representaban patrones temporales, como la hora del día, el día de la semana y día del mes (condicionales para VAE).

Los datasets producidos por ingeniería de características, son de periodicidad horaria y están en un rango de fechas común con el dataset de periodicidad horaria de EUR/USD. En cuanto a la predicción por horas se usan las características tal cual como se generan desde el extractor. Para manejo de datos diarios se estiman 24 horas del dataset generado.

9. PREPARACIÓN DE LOS DATOS

En esta sección se describen los pasos realizados para preparar los datos antes de utilizarlos en el entrenamiento de los modelos predictivos. La preparación incluyó la limpieza, transformación, integración y la construcción de datasets específicos para los experimentos realizados. Estas tareas fueron fundamentales para garantizar que los datos estuvieran en un formato adecuado para los modelos de aprendizaje profundo.

9.1. PREPROCESAMIENTO DE INFORMACIÓN

Se aplicaron técnicas estándar de alineamiento por fecha, división en training/test y normalización:

- **Alineación por Fecha:** dado que se utilizó múltiples datasets, se buscó el rango de fechas común entre ellos y se eliminaron los registros fuera de este rango común. Obteniendo datasets procesados de 93.084 filas desde 2005 hasta 2020.
- **Valores Faltantes:** En las series de tiempo de Forex, no se incluyen los fines de semana, no se utilizó interpolación u otras técnicas de imputación, debido a que la imputación de los dos días por semana equivale a introducir $2/7 = 28.5\%$ de datos inexistentes, que podría hacer que los modelos de Deep Learning puedan aprender patrones no deseados de los datos imputados para el fin de semana. La estrategia utilizada fue ignorar los fines de semana para tomar datos sólo en días hábiles.
- **División de Datasets para Entrenamiento y Validación:** Los datasets para cada experimento se dividieron en 3 (no en dos), uno se usó para entrenar tanto el autoencoder para el extractor de características como los sistemas de generación de datos sintéticos (20%), el segundo se usó para para el entrenamiento de todos los modelos predictivo (60%) y el tercero para validación de los predictores (20%). Se usó un dataset más grande que los otros para entrenamiento (60%) porque un experimento consiste en variar la cantidad de datos desde 3 meses hasta 4 años, es decir requiere 25.960 registros, mientras que para validación solo se usó un año en todos los casos. Se usa un dataset separado para el generador de datos sintéticos y el extractor de características porque de usar el mismo dataset que para el entrenamiento de los predictores se puede introducir overfitting, haciendo que el error en entrenamiento sea engañosamente menor respecto a usar datasets separados.
- **Límite Configurable de Número de Registros para Entrenamiento:** Todos los modelos incluyendo el predictores, autoencoder y sistemas de generación de datos sintéticos tienen un parámetro llamado `max_steps` que permite configurar exactamente cuántos registros del dataset de entrada usar para el entrenamiento, esto permite usar un solo dataset de entrenamiento para los experimentos en lugar de generar uno para cada experimento con el número de registros requerido. En este caso necesariamente se usan los datos más antiguos (desde el 2005) para hacer entrenamiento y datos más recientes para hacer validación, respetando siempre el orden y la temporalidad de las señales para que los diferentes modelos puedan captar las señales de la mejor manera. Las series de tiempo financieras, como el tipo de cambio EURO/USD, suelen tener una fuerte dependencia temporal. Los precios de hoy están influenciados por los precios de ayer y los días anteriores.

- **Normalización:** Todos los valores numéricos fueron **escalados entre 0 y 1** utilizando el método Min-Max, ajustando el escalador en el dataset de entrenamiento, únicamente y guardando los parámetros de normalización como un archivo JSON para realizar la **desnormalización** de las predicciones para su uso en la estrategia heurística de trading. A pesar de la normalización entre 0 y 1, no se puede garantizar que existan valores negativos en el dataset, debido a que el dataset de validación puede contener datos menores a los mínimos usados en la normalización. Por tanto, los modelos usados tienen **la función de activación de entrada Linear, y TanH (Tangente Hiperbólica) para capas intermedias**, debido a que funciones que truncan entradas negativas como **ReLU** pueden introducir error por pérdida de información si una entrada negativa no coincide con un peso negativo durante la inicialización aleatoria con distribución uniforme de pesos usada y si el learning-rate es lo suficientemente pequeño como los usados en los experimentos, este problema no se corrige durante el entrenamiento.

10. PREPARACIÓN DE DATASETS ESPECÍFICOS PARA CADA MODELO PREDICTIVO

Los datasets se prepararon según los requisitos específicos de los modelos predictivos, esta preparación fue hecha por el módulo predictor al cargar el plugin correspondiente a cada modelo antes de realizar el entrenamiento o inferencia del modelo entrenado en el dataset de validación.

10.1. PREDICTORES CON ANN, LSTM Y CONV1D

Para estos modelos se aplicaron ventanas deslizantes (sliding windows) de 288 ticks u observaciones horarias, es decir aproximadamente dos semanas, para generar matrices tridimensionales, donde cada fila representa una ventana de tiempo consecutiva.

Estas ventanas deslizantes permitieron capturar relaciones temporales en las series para los modelos basados en convoluciones y redes densas.

10.2. PREDICTORES CON TRANSFORMERS

Para el modelo basado en Transformers, se adicionaron columnas de codificación posicional (Positional Encoding), que permitieron a los Transformers distinguir entre posiciones relativas dentro de las secuencias temporales. Estas codificaciones se calcularon utilizando funciones seno con frecuencia variable para cada característica de entrada. En este caso también se utiliza un sliding window de 288 ticks, por lo tanto, se tiene una matriz tridimensional.

11. MODELADO

En esta sección se describen las arquitecturas y configuraciones de los modelos predictivos utilizados en el proyecto, así como el diseño experimental para evaluar su desempeño. El objetivo del modelado fue explorar cómo diferentes arquitecturas de aprendizaje profundo (ANN, Conv1D, LSTM y Transformers) y combinaciones de datos impactan la precisión y robustez de las predicciones de la tasa de cambio EUR/USD. Además, se detallan las técnicas de validación, división de los datos y los hiperparámetros seleccionados para optimizar los modelos.

11.1. MODELOS DE DEEP LEARNING E HIPERPARÁMETROS

Los modelos de Deep Learning fueron implementados usando la librería **Keras** que hace parte de Tensorflow desde la versión 2.0, esta librería permite ejecutar tanto el entrenamiento, como la inferencia usando GPU, lo que es ideal dado el número de experimentos realizados.

Entre los hiperparámetros clave se encuentra la tasa de aprendizaje (**learning_rate**), inicialmente establecida en 1E-3, la cual se ajustó dinámicamente durante el entrenamiento mediante la técnica **ReduceLROnPlateau de Keras**. Este mecanismo reduce la tasa de aprendizaje cuando no se observa una disminución significativa en la función de pérdida (loss) del conjunto de validación durante un período de 25 épocas (ciclos completos de entrenamiento sobre todo el dataset).

El número máximo de épocas se fijó en 400, pero se implementó un mecanismo de **Early Stopping** con una paciencia de 100 épocas. Esto detiene el entrenamiento prematuramente si no hay mejora en la pérdida del MAE de validación durante este período, **previniendo el sobreajuste**. La función de pérdida (loss) utilizada fue una combinación aditiva de la función Huber (híbrido entre el error cuadrático medio MSE para errores pequeños y el error absoluto medio MAE para errores grandes) y el **Maximum Mean Discrepancy (MMD)**, ponderado por un factor Sigma de 0.01.

El optimizador seleccionado fue Adam con sus parámetros predeterminados. Las arquitecturas de las capas de los modelos variaron en tamaño, como se detalla en las visualizaciones de los modelos.

Se empleó la función de **activación lineal** en la primera y la última capa de todos los modelos para permitir rangos de valores amplios en la entrada y la salida. Para las capas internas, se utilizó la función de **activación tanh y ReLu**. Se aplicó **regularización L2** con un factor l2_reg de 1E-5 para **mitigar el sobreajuste**, complementado eficazmente por el Early Stopping. En las capas convolucionales, se utilizó un stride de 2 para reducir la dimensión temporal. Las capas de entrada para todos los tipos de modelos fueron tridimensionales, con una estructura de (batch_size, window_size, num_features), lo que requirió un procesamiento paralelo de las características para las redes neuronales artificiales (ANN) debido a la naturaleza bidimensional de sus capas densas.

El tamaño de la ventana deslizante (sliding window) utilizada para generar las secuencias de entrada fue de 288 ticks para modelos de largo plazo y 144 ticks para modelos de corto plazo, con un tamaño de lote (batch size) de 128.

Es importante destacar que, mientras que los hiperparámetros de los modelos de predicción se ajustaron manualmente, los parámetros de la estrategia de trading se optimizaron mediante un

método automático de búsqueda utilizando algoritmos genéticos a través de la librería **DEAP**. En futuros trabajos, se planea incorporar la optimización de los hiperparámetros de entrenamiento de los predictores también con algoritmos genéticos.

Se hicieron predicciones para largo y corto plazo, dado que estas predicciones serán usadas en una estrategia heurística de trading (sección 5.7) que usa un conjunto de predicciones de corto y largo plazo para decidir la dirección (Short o Long), ejecución y cierre de órdenes de trading de EURUSD usando una simulación de un broker de Forex con la librería **backtrader**.

Para las predicciones de corto plazo, se usaron los horizontes predictivos de 1 hora, 2h, 3h, 4h, 5h y 6h; mientras que para las predicciones de largo plazo se utilizaron los horizontes predictivos de 24 horas, 48h, 72h, 96h, 120h, 144h.

En pruebas manuales preliminares realizadas con diversas combinaciones de modelos, se determinó que el mejor desempeño predictivo se logra al adicionar una capa bidireccional **LSTM y una capa Bayesiana** al final de todos los modelos. En todas las arquitecturas usadas, se pueden diferenciar dos secciones, **el extractor de características** compuesto por las primeras 2 capas (o módulos de capas) y los **cabezales predictivos para Direct Multi-Output** que se entrenan en paralelo y toman los mismos datos de salida del extractor de características.

11.2. MODELO DE ARTIFICIAL NEURAL NETWORKS (ANN) CON CAPAS DENSAS

La red neuronal artificial (ANN) diseñada para realizar predicciones multivariadas está compuesta por una capa de entrada seguida de varias capas densas (fully connected), lo que permite capturar patrones complejos en los datos. Utiliza funciones de activación *Tanh* para introducir no linealidades y una función lineal en la salida para producir valores continuos.

La arquitectura implementada consta de dos capas densas consecutivas con 24 neuronas que actúan como extractor de características común para los predictores que se entrenan en paralelo (Direct Multi-Output) que procesan en paralelo cada una de las características, seguidas por una capa de *concatenación*, que alimenta los módulos llamados cabezales predictores compuestos por dos capas densas seguidas por una capa LSTM y una capa Bayesiana de salida.

Esta estructura responde a una estrategia común de reducción progresiva de dimensionalidad, que permite al modelo abstraer representaciones de mayor nivel a partir de los datos de entrada. Al usar **activación Tanh**, en las capas internas y **activación Lineal** para las capas de entrada y salida se permite el uso de entradas con valores por fuera del rango de normalización usado mientras facilita la modelación de relaciones no lineales propias de series temporales financieras como el tipo de cambio EUR/USD.

Las dos primeras capas desde la entrada fueron llamadas **extractor de características**, la salida de estas capas es la entrada común para un sistema de predicción tipo Direct Multi-Output, donde estos extractores de características producen los datos para los **cabezales predictivos** que se entrenan por separado en paralelo para cada horizonte predictivo en horas.

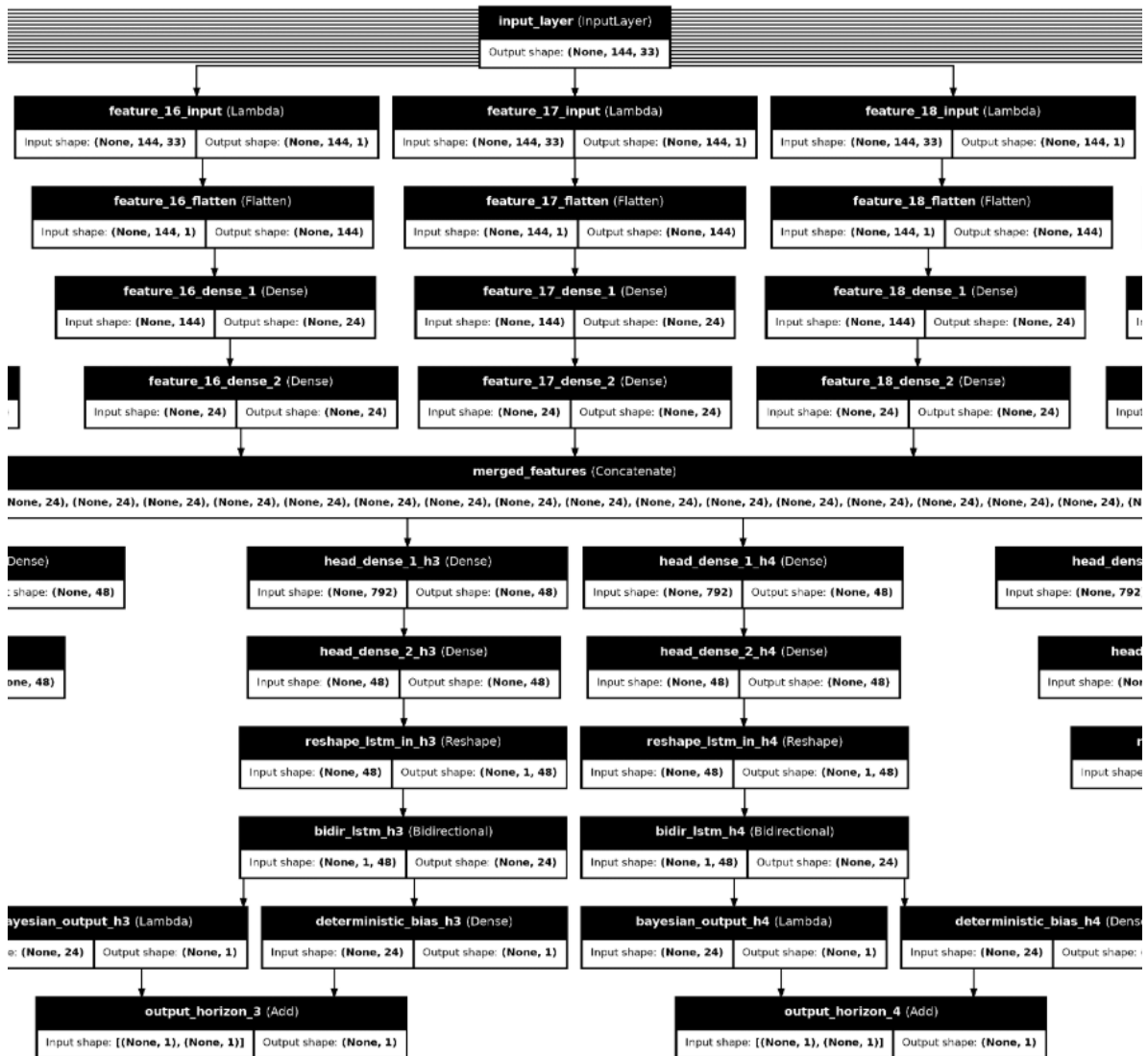


Figura 6 – Modelo de capas densas ANN con Parallel Branch Feature Processing y Direct Multi-Output (Fragmento de plot generado con `keras.utils.plot_model`)

Como se observa en la Figura 6, se muestra un fragmento del extractor de características para el modelo ANN (el diagrama del modelo completo se encuentra en el **Anexo A**), se usaron ramas paralelas partiendo de la capa de entrada para procesar cada característica por separado con dos capas dense, ya que las capas densas tienen entrada y salida bidimensional (`batch_size`, `num_features`) y se requiere que su concatenación sea tridimensional para que sea compatible con los experimentos siguientes que usan una capa inicial Conv1D para el predictor que recibe una entrada tridimensional. Esta técnica también permitió obtener mejores predicciones que mezclar todas las características para toda la ventana deslizante usando capas densas comunes a todas las entradas.

También se observa en la Figura 6, que se usó Direct Multi-Output, donde las capas concatenadas iniciales se alimentan a módulos de capas de salida que terminan en una **capa LSTM** seguida por una capa Bayesiana implementada con Tensorflow-Probability 0.25, esta capa bayesiana tiene como bias una capa densa.

La **capa Bayesiana** usa un número de estimaciones (25) usando inicializaciones con variaciones aleatorias de una capa densa, variaciones determinadas por la probabilidad a priori dado el historial de valores de entrada para calcular la salida de capa la bayesiana (**predicción**) como la **media de estas estimaciones y la desviación estándar como la incertidumbre de la predicción** para cada punto.

11.3. MODELO DE RED NEURAL CONVOLUCIONAL UNIDIMENSIONAL (CNN)

El modelo basado en redes neuronales convolucionales (CNN) está diseñado para procesar series temporales, combinando capas convolucionales unidimensionales.

La arquitectura comienza con un extractor de características compuesto por dos capas **Conv1D de 48 filtros** con $\text{stride}=2$ para reducir la dimensión temporal a la mitad en cada capa. Estas capas convolucionales tienen activación *Tanh* para capas internas y *Linear* para capas de Entrada y salida.

Al igual que en el modelo ANN, se usa **Direct Multi-Output** para entrenar los predictores de 6 horizontes predictivos diferentes a la vez usando el mismo **extractor de características** (dos primeras capas Conv1D) compartido entre todos los cabezales predictivos. El uso del Parallel Branch Feature Processing usado en el modelo ANN (sección 11.2) no hace falta en este modelo, ya que las capas Conv1D ya reciben una entrada tridimensional (batch_size , window_size , num_features) y no se pierde la información temporal (posición de un dato en ventana deslizante de entrada), como ocurre con las capas Dense de entrada bidimensional (batch_size , num_features).

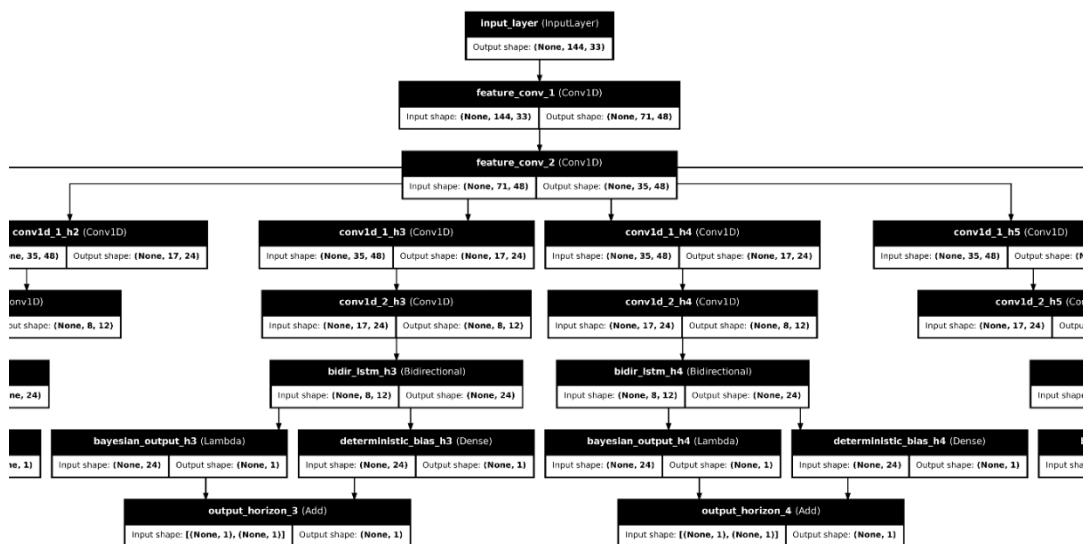


Figura 7 – Modelo CNN con Direct Multi-Output (Fragmento de plot generado con `keras.utils.plot_model`)

Como se observa en la Figura 7, se observa un fragmento del modelo CNN (el diagrama del modelo completo se encuentra en el **Anexo B**) con los cabezales predictivos se componen de dos capas Conv1D seguidos por una capa LSTM y la capa Bayesiana de salida con capa de bias tipo Dense.

11.4. MODELO LSTM

El modelo basado en redes LSTM (Long Short-Term Memory) está diseñado específicamente para capturar dependencias temporales en secuencias. Utiliza capas recurrentes con activación *Tanh* para modelar dinámicas complejas en los datos, y finaliza con una capa de salida lineal que permite realizar predicciones continuas multi-horizonte.

La arquitectura comienza con un extractor de características compuesto por dos capas **LSTM de 48 filtros** con capas intermedias de **AveragePooling1D** para reducir la dimensión temporal a la mitad en cada capa. Estas capas LSTM tienen activación *Tanh* para capas internas y *Linear* para capas de Entrada y salida. En el plot del modelo mostrado en la Figura 8, se muestra el tamaño de salida de las capas como 96 en lugar de 48, esto es porque se usaron capas LSTM Bidireccionales, que buscan patrones temporales en ambas direcciones dentro de su ventana de datos (sequence), produciendo el número de filtros como salida para el orden directo y el inverso y por tanto produciendo el doble de salidas que el número de filtros usados por capa.

Al igual que en el modelo ANN, se usa **Direct Multi-Output** para entrenar los predictores de 6 horizontes predictivos diferentes a la vez usando el mismo **extractor de características** (dos primeras capas LSTM) compartido entre todos los cabezales predictivos. El uso del Parallel Branch Feature Processing usado en el modelo Dense no hace falta en este modelo, ya que las capas Conv1D ya reciben una entrada tridimensional (batch_size, sequence_size, num_features) y no se pierde la información temporal (posición de un dato en ventana deslizante de entrada), como ocurre con las capas Dense de entrada bidimensional (batch_size, num_features).

El uso de una capa bidireccional LSTM mejora marginalmente el desempeño respecto a una capa unidireccional debido a que mientras en una capa unidireccional solamente se buscan patrones con LSTM en una dirección en su ventana deslizante de tamaño `sequence_len`, con la capa bidireccional el LSTM busca patrones en ambas direcciones en la ventana deslizante de entrada, encontrando algunos patrones adicionales en los datos pasados, que causan su mejora en desempeño, sin violar la causalidad, ya que su entrada tridimensional de tamaño (batch_size, sequence_len, feature_num) solo contiene datos pasados.

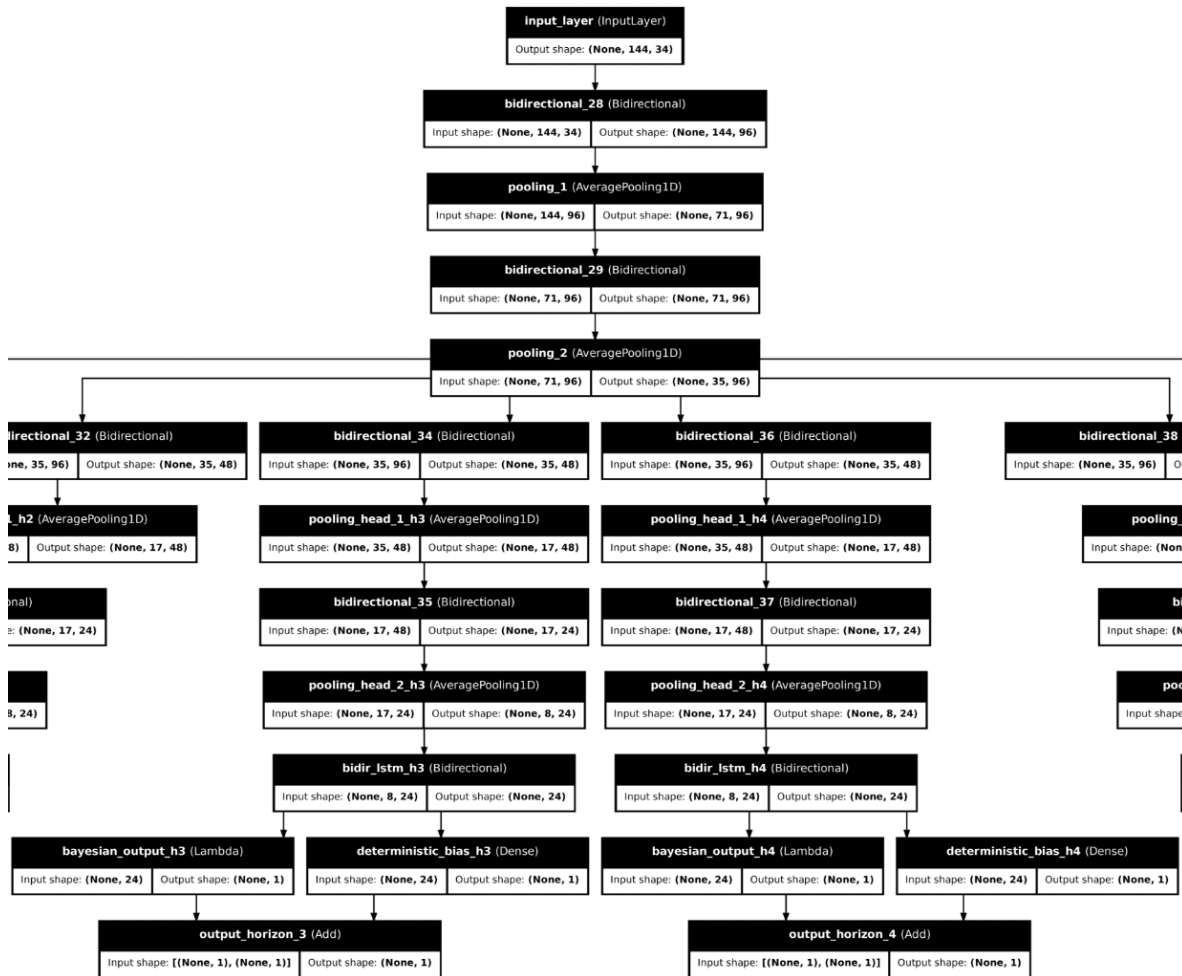


Figura 8 – Modelo Bidir-LSTM con Direct Multi-Output (Fragmento de plot generado con `keras.utils.plot_model`)

En la Figura 8 se puede apreciar un fragmento del extractor de características del modelo LSTM (el diagrama del modelo completo se encuentra en el **Anexo C**), que cuenta con dos capas LSTM mientras los cabezales predictivos usan tres capas LSTM seguidas por la capa bayesiana de salida.

11.5. MODELO BASADO EN TRANSFORMERS

El modelo basado en Transformers aplica mecanismos de atención para capturar relaciones complejas y de largo alcance en series temporales. Combina capas densas con módulos de autoatención para transformar dinámicamente la representación de los datos, finalizando en una capa de salida lineal que permite predicciones continuas multivariadas

La arquitectura usada para Transformers, después de la capa de entrada, tiene como **extractor de características**, dos módulos consecutivos de **transformers** compuestos por capas densas y módulo de autoatención (capa aditiva de la anterior más MultiHeadAttention de 2 cabezales, proyección y reshape), de tamaño 64 y tienen los siguientes componentes, por ejemplo, para la capa 1:

- **Reformateo para Autoatención (reshape_1):** La salida (None, 64) se convierte en (None, 8, 8), lo que indica 8 vectores de 8 dimensiones.
- **Autoatención Multi-Cabecal (mha_layer_1):** Aplica atención sobre los vectores de 8 dimensiones. Calcula pesos de importancia entre cada vector.
- **Proyección densa posterior (mha_projection_1):** Transforma la salida de la autoatención con una capa Dense con Tanh, ajustando la escala de los valores.
- **Reformateo Inverso (reshape_back_1):** Convierte la salida de (None, 8, 8) de vuelta a (None, 64).
- **Suma Residual (residual_add_1):** Se suma la entrada original (None, 64) a la salida procesada para preservar información inicial junto con los datos con atención aplicada.

El **extractor de características** proporciona entradas comunes para todos los cabezales predictivos de la arquitectura Direct Multi-Output, estos **cabezales predictivos** están compuestos por dos módulos de transformers, una capa LSTM y una capa Bayesiana de salida.

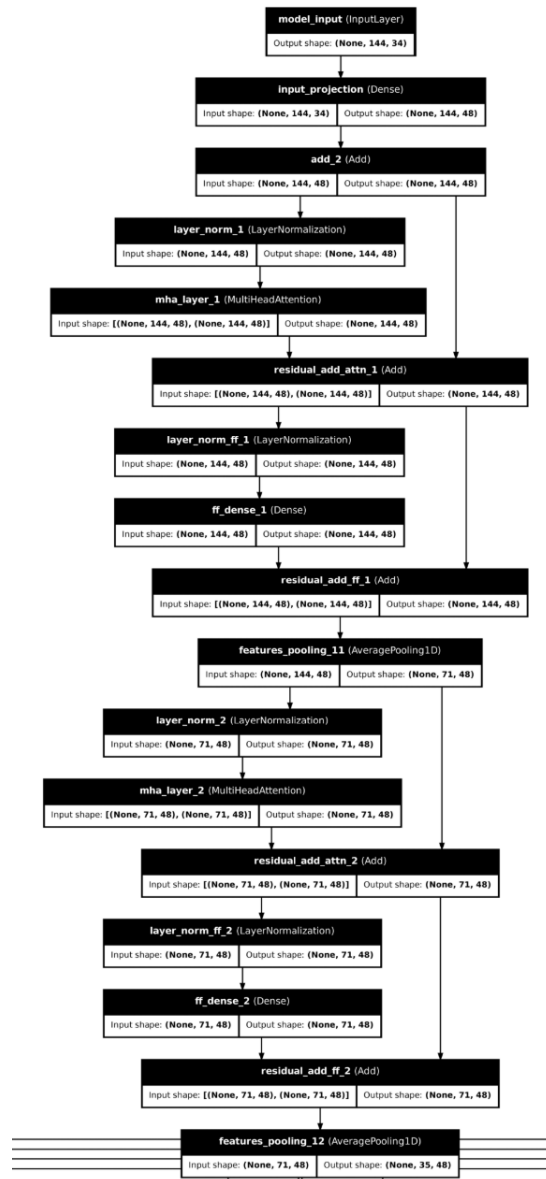


Figura 9 – Extractor de características de Modelo de Transformers con Direct Multi-Output (Fragmento de plot generado con `keras.utils.plot_model`)

En la Figura 9 se puede apreciar un fragmento del extractor de características del modelo Transformers (el diagrama del modelo completo se encuentra en el **Anexo D**), dado que el modelo completo incluyendo los cabezales predictivos es demasiado grande para mostrar en un plot con texto legible en una sola hoja. Los cabezales predictivos de Direct Multi-Output tienen la misma estructura de dos módulos de transformers más una capa LSTM y una capa bayesiana de salida similar a los modelos anteriores.

12. DESARROLLO DE EXPERIMENTOS

En esta sección se desarrollaron varios conjuntos de experimentos cada uno para cumplir un objetivo específico.

12.1. EXPERIMENTOS DE VARIACIÓN DE LA CANTIDAD DE REGISTROS DE ENTRENAMIENTO

En este experimento se buscó evaluar de manera sistemática cómo la variación en la cantidad de registros de entrenamiento impacta el desempeño de diferentes arquitecturas de Deep Learning, específicamente Redes Neuronales Artificiales (ANN), Redes Neuronales Convolucionales (CNN), Redes de Memoria a Corto-Largo Plazo (LSTM) y modelos Transformer, en la tarea de predecir la tasa de cambio EUR/USD. Para este análisis, los modelos se entrenaron utilizando un conjunto de 8 características básicas, que incluyen los precios de apertura(O), máximo(H), mínimo(L) y cierre(C), junto con variables derivadas directamente de estos valores (C-O, H-L,H-O,O-L).

El desempeño de cada modelo, entrenado con distintas magnitudes de datos históricos (que varían desde el equivalente a tres meses hasta cuatro años de observaciones), se cuantificó mediante métricas de error estándar en problemas de regresión, como el Error Absoluto Medio (MAE) y el Coeficiente de Determinación (R^2). Estas evaluaciones se realizaron tanto para predicciones con granularidad horaria a corto plazo (horizonte predictivo de 1 a 6 horas) como a largo plazo (horizonte predictivo de 24 a 144 horas equivalente a 1-6 días), permitiendo así comprender la sensibilidad del desempeño predictivo a la cantidad de datos en estos horizontes temporales.

12.1.1. DESCRIPCIÓN DE LOS DATOS

Para este conjunto de experimentos se usó un número variable de registros en el dataset de entrenamiento, todas las periodicidades son de una hora. Las características usadas fueron:

- **Close:** Valor de cierre en el periodo, también es el valor que se va a predecir.
- **Open:** Valor de apertura en el periodo, también es el Close del periodo anterior
- **High:** Valor más alto durante el periodo
- **Low:** Valor más bajo durante el periodo
- **High-Low:** Variación máxima durante el periodo
- **Close-Open:** Variación entre periodos
- **High-Open:** Máxima subida en el periodo relativo al precio de Open
- **Open-Low:** Máxima bajada en el periodo relativo al precio de Open

El número de registros usados en las predicciones realizadas para este experimento fueron:

- **Tres meses:** 1.575 registros
- **Seis meses:** 3.150 registros
- **Un año:** 6.300 registros
- **Dos años:** 12.600 registros

- **Cuatro años:** 25.200 registros

Antes de describir los resultados de los modelos de redes neuronales (ANN, Transformers, LSTM, CNN LARGO PLAZO) aplicados a la predicción de la tasa de cambio euro/dólar estadounidense, de describirán brevemente las técnicas de descomposición de series de tiempo utilizadas: STL, Wavelet y Multitaper.

En nuestro caso, al utilizar el wavelet db4, que tiene un ancho de 5 ticks, el cálculo original para un punto dado incluiría información de 2 ticks en el pasado, el punto actual y 2 ticks en el futuro. Para evitar el uso de información futura en nuestro contexto de predicción, se realizó un desplazamiento del valor calculado de forma que cada valor transformado por wavelet utiliza una ventana de 5 ticks pasados únicamente.

A continuación, se describirán los resultados mediante el análisis de las métricas MAE y R^2 al aplicar los modelos de redes neuronales (ANN, Transformers, LSTM, CNN LARGO PLAZO) a la serie de tiempo de la tasa de cambio euro/dólar estadounidense y también después de aplicar alguna de las técnicas de descomposición para preprocesar los datos.

12.1.2. ANÁLISIS DEL COMPORTAMIENTO DEL MAE PARA PREDICCIONES EN CORTO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Test	Variación Test - Validación
ann	1575	Serie de tiempo	0,002584	0,004025	0,002347	55,77%	41,69%
ann	3150	Serie de tiempo	0,002479	0,001966	0,002128	20,69%	8,24%
ann	6300	Serie de tiempo	0,002326	0,001858	0,002113	20,12%	13,72%
ann	12600	Serie de tiempo	0,000308	0,000265	0,000372	13,96%	40,38%
ann	25200	Serie de tiempo	0,000138	0,000117	0,000172	15,22%	47,01%
cnn	1575	Serie de tiempo	0,002561	0,002682	0,004483	4,72%	67,15%
cnn	3150	Serie de tiempo	0,001843	0,001535	0,002098	16,71%	36,68%
cnn	6300	Serie de tiempo	0,001051	0,000855	0,001269	18,65%	48,42%
cnn	12600	Serie de tiempo	0,000158	0,000124	0,000205	21,52%	65,32%
cnn	25200	Serie de tiempo	0,000137	0,000108	0,000134	21,17%	24,07%
lstm	1575	Serie de tiempo	0,002421	0,002455	0,003762	1,40%	53,24%
lstm	3150	Serie de tiempo	0,002220	0,001794	0,003018	19,19%	68,23%
lstm	6300	Serie de tiempo	0,001967	0,001637	0,002278	16,78%	39,16%
lstm	12600	Serie de tiempo	0,001762	0,001509	0,001757	14,36%	16,43%
lstm	25200	Serie de tiempo	0,000851	0,000552	0,000669	35,14%	21,20%
transformer	1575	Serie de tiempo	0,002561	0,002190	0,002754	14,49%	25,75%
transformer	3150	Serie de tiempo	0,002472	0,002039	0,002912	17,52%	42,82%
transformer	6300	Serie de tiempo	0,002322	0,001865	0,002285	19,68%	22,52%
transformer	12600	Serie de tiempo	0,002076	0,001802	0,002082	13,20%	15,54%
transformer	25200	Serie de tiempo	0,001011	0,000654	0,000738	35,31%	12,84%

Tabla 4 - Resultados MAE de experimentos de variación de número de registros de entrenamiento para corto plazo

En la Tabla 4, los resultados del Error Absoluto Medio (MAE) sobre el conjunto de validación revelan una tendencia general de mejora en el desempeño a medida que se incrementa la cantidad de registros de entrenamiento. Esta disminución del error sugiere que un mayor volumen de datos permite a los modelos capturar patrones más robustos y generalizables. El modelo CNN, en

particular, muestra una mejora continua y significativa, reduciendo su MAE de manera progresiva al aumentar el tamaño del conjunto de datos, lo que evidencia una eficiente capitalización de la información adicional.

Por otra parte, la reducción del error no se presenta de forma lineal en todas las arquitecturas. El modelo ANN, por ejemplo, exhibe una mejora importante al pasar de 6,300 a 12,600 registros, pero el ritmo de mejora se modera en incrementos posteriores de datos. Este comportamiento es indicativo de un fenómeno de rendimientos decrecientes, donde la ganancia en precisión es progresivamente menor por cada nuevo bloque de datos añadido, una vez que los patrones más relevantes de la serie han sido asimilados por el modelo.

Al realizar una comparación directa entre los modelos con los volúmenes de datos más altos (12,600 y 25,200 registros), la arquitectura CNN se consolida como la de mayor precisión, logrando el MAE de validación más bajo en ambos casos. Le sigue de cerca el modelo ANN, que también alcanza un nivel de error muy competitivo. Por su parte, los modelos LSTM y Transformer, aunque muestran una clara tendencia a la mejora, no alcanzan los mismos niveles de exactitud que CNN y ANN con las mayores cantidades de datos disponibles para este experimento.

Las columnas "Variación Train - Test" y "Variación Test - Validación" revelan la diferencia en el rendimiento entre las diferentes etapas de evaluación. Valores altos en estas columnas indican un posible sobreajuste (buen rendimiento en entrenamiento, pero peor en prueba) o una falta de robustez en la generalización a datos no vistos. Por ejemplo, algunos modelos CNN muestran variaciones muy altas, lo que sugiere que, aunque puedan ajustarse bien a los datos de entrenamiento, su rendimiento se degrada significativamente con datos nuevos. En contraste, modelos como el LSTM en algunas configuraciones muestran variaciones más bajas, lo que podría indicar una mejor capacidad de generalización. Es importante notar que la "Variación Test - Validación" también es relevante para evaluar la estabilidad del modelo ante nuevos datos no utilizados durante el ajuste de hiperparámetros.

12.1.3. ANÁLISIS DEL COMPORTAMIENTO DEL R2 PARA CORTO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Test	Variación Test - Validación
ann	1575	Serie de tiempo	0,977502	0,960747	0,998815	1,71%	3,96%
ann	3150	Serie de tiempo	0,972100	0,994315	0,998940	2,29%	0,47%
ann	6300	Serie de tiempo	0,977669	0,995788	0,998945	1,85%	0,32%
ann	12600	Serie de tiempo	0,999913	0,999909	0,999964	0,00%	0,01%
ann	25200	Serie de tiempo	0,999995	0,999981	0,999987	0,00%	0,00%
cnn	1575	Serie de tiempo	0,977952	0,990500	0,995329	1,28%	0,49%
cnn	3150	Serie de tiempo	0,984364	0,997241	0,999015	1,31%	0,18%
cnn	6300	Serie de tiempo	0,994707	0,998970	0,999634	0,43%	0,07%
cnn	12600	Serie de tiempo	0,999969	0,999977	0,999988	0,00%	0,00%
cnn	25200	Serie de tiempo	0,999996	0,999985	0,999994	0,00%	0,00%
lstm	1575	Serie de tiempo	0,979796	0,993058	0,997176	1,35%	0,41%
lstm	3150	Serie de tiempo	0,978990	0,996372	0,998345	1,78%	0,20%
lstm	6300	Serie de tiempo	0,985470	0,996997	0,998997	1,17%	0,20%
lstm	12600	Serie de tiempo	0,997674	0,997434	0,999336	0,02%	0,19%
lstm	25200	Serie de tiempo	0,999881	0,999673	0,999906	0,02%	0,02%
transformer	1575	Serie de tiempo	0,977671	0,994581	0,998278	1,73%	0,37%
transformer	3150	Serie de tiempo	0,972019	0,994962	0,997258	2,36%	0,23%
transformer	6300	Serie de tiempo	0,977729	0,995801	0,998868	1,85%	0,31%
transformer	12600	Serie de tiempo	0,996384	0,996060	0,998987	0,03%	0,29%
transformer	25200	Serie de tiempo	0,999790	0,999443	0,999862	0,03%	0,04%

Tabla 5 - Resultados R2 de experimentos de variación de número de registros de entrenamiento para corto plazo

En la Tabla 5 se puede evidenciar que a medida que se incrementa la cantidad de datos de entrenamiento, el R2 en los conjuntos de prueba y validación tiende a aumentar, acercándose a 1. Los modelos son capaces de capturar mejor la varianza de la tasa de cambio con más información. Sin embargo, se observa en algunos casos, el incremento en el R2 se vuelve menos significativo al añadir grandes cantidades de datos (por ejemplo, al pasar de 12600 a 25200 muestras), lo que podría indicar un rendimiento marginal decreciente en el rendimiento del modelo, entendiendo que el máximo coeficiente para el R2 es de 1.

Se observan, también, diferencias en los valores de R2 entre los distintos tipos de modelos para una misma cantidad de datos. Por ejemplo, para cantidades de datos más grandes, el modelo Transformer a menudo muestra un R2 en el conjunto de prueba relativamente alto en comparación con otros modelos, sugiriendo una mejor capacidad para explicar la varianza de los datos no vistos. Por otro lado, el modelo CNN, en algunas instancias, muestra una mayor discrepancia entre el R2 de entrenamiento y los conjuntos de prueba/validación (mayor sobreajuste), lo que implica una menor capacidad de generalización a nuevos datos.

Valores negativos grandes en las columnas de variación podrían llegar a indicar un sobreajuste significativo, donde el modelo se ajusta muy bien a los datos de entrenamiento, pero pierde precisión con datos nuevos.

12.1.4. EVOLUCIÓN DEL MAE AL AUMENTAR LA CANTIDAD DE REGISTROS

La Figura 10 muestra la variación del resultado del MAE con respecto a la cantidad de datos con los que cada uno de los modelos (ANN, CNN, LSTM y Transformer) se va entrenando. Los resultados de la visualización corresponden al MAE de validación.

Métricas de Desempeño (MAE) al aumentar la cantidad de datos de entrenamiento

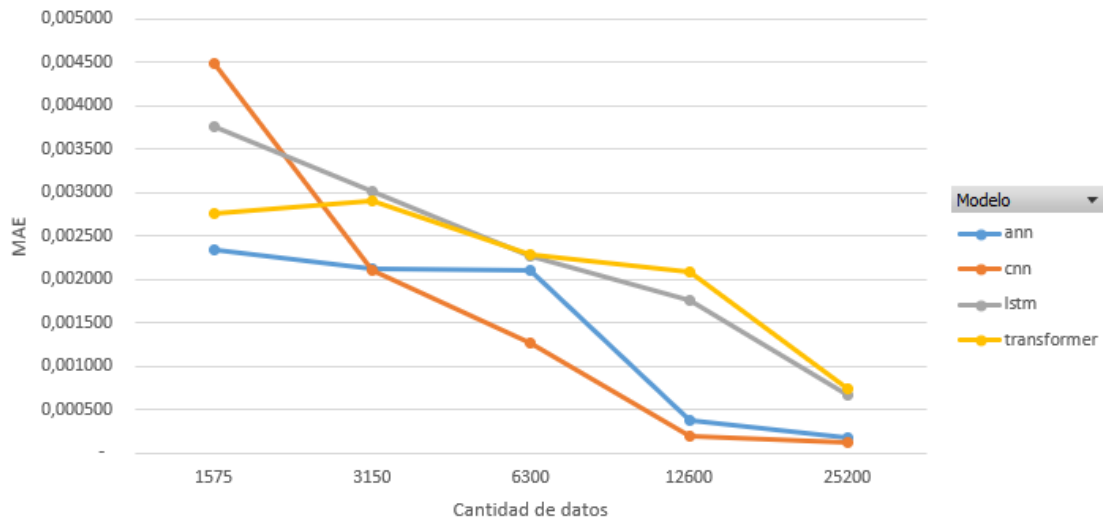


Figura 10 - Desempeño de MAE al aumentar la cantidad de datos - corto plazo

Como lo muestra la Figura 10 todos los modelos tienen una relación inversamente proporcional con respecto a la cantidad de datos usados para el entrenamiento y el MAE de validación para los mismos. El modelo que se ve mayormente beneficiado por la cantidad de datos insertada en el proceso de entrenamiento es el de CNN, el cual inicia con el mayor error cuando se utilizan sólo 1.157 ticks y al usar 25.200 (la mayor cantidad de datos) es el que menor error presenta.

12.1.5. MEJOR PREDICCIÓN LOGRADA EN LOS MODELOS ENTRENADOS PARA CORTO PLAZO

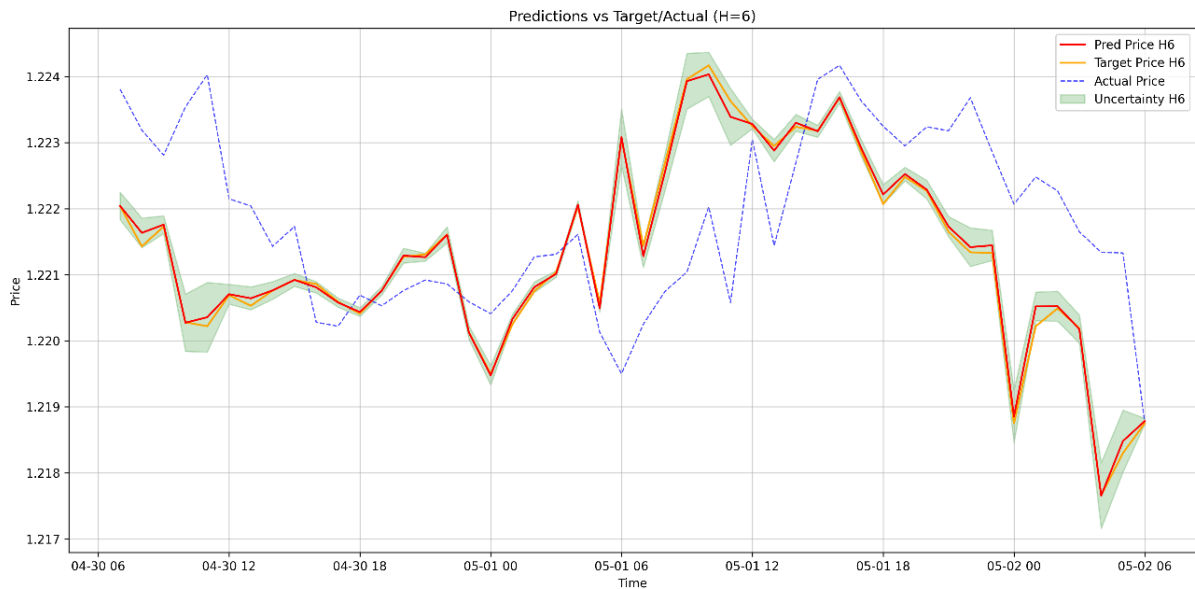


Figura 11 - Predicción para CNN con 25.200 steps con periodos de corto plazo

Se encuentra que la mejor predicción en esta etapa la realiza el modelo CNN entrenado con 25.200 steps para periodos de una hora. Como se puede ver en la Figura 11 la serie temporal generada por el modelo muestra similitud con el comportamiento de la señal real. Lo cual corresponde con las métricas de evaluación usadas, ya que se posee un MAE (Error Absoluto Medio) bajo y un R2 (Coeficiente de Determinación) alto, cercano a 1, que señala una buena capacidad del modelo para explicar la varianza de la serie de tiempo real.

12.1.6. ANÁLISIS DEL COMPORTAMIENTO DE MAE PARA LARGO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Test	Variación Test - Validación
ann	1575	Serie de tiempo	0,011289	0,012831	0,016243	13,66%	26,59%
ann	3150	Serie de tiempo	0,002471	0,006551	0,011141	165,12%	70,07%
ann	6300	Serie de tiempo	0,004131	0,004902	0,008098	18,66%	65,20%
ann	12600	Serie de tiempo	0,004432	0,004248	0,005209	4,15%	22,62%
ann	25200	Serie de tiempo	0,005467	0,004331	0,005175	20,78%	19,49%
cnn	1575	Serie de tiempo	0,006493	0,009193	0,025143	41,58%	173,50%
cnn	3150	Serie de tiempo	0,005906	0,007036	0,013144	19,13%	86,81%
cnn	6300	Serie de tiempo	0,005773	0,005732	0,017687	0,71%	208,57%
cnn	12600	Serie de tiempo	0,004498	0,004496	0,005953	0,04%	32,41%
cnn	25200	Serie de tiempo	0,005538	0,004081	0,005282	26,31%	29,43%
lstm	1575	Serie de tiempo	0,005658	0,017276	0,025161	205,34%	45,64%
lstm	3150	Serie de tiempo	0,003959	0,025109	0,028815	534,23%	14,76%
lstm	6300	Serie de tiempo	0,004550	0,015311	0,020129	236,51%	31,47%
lstm	12600	Serie de tiempo	0,004587	0,005457	0,007348	18,97%	34,65%
lstm	25200	Serie de tiempo	0,004586	0,004983	0,006235	8,66%	25,13%
transformer	1575	Serie de tiempo	0,011182	0,011294	0,012601	1,00%	11,57%
transformer	3150	Serie de tiempo	0,008815	0,011224	0,013871	27,33%	23,58%
transformer	6300	Serie de tiempo	0,004995	0,007036	0,018868	40,86%	168,16%
transformer	12600	Serie de tiempo	0,004919	0,004786	0,007131	2,70%	49,00%
transformer	25200	Serie de tiempo	0,007547	0,004874	0,006788	35,42%	39,27%

Tabla 6 - Resultados MAE de experimentos de variación de número de registros de entrenamiento para largo plazo

El MAE tanto en el conjunto de prueba como en el de validación, al agregar más datos, tiende a disminuir. Esto sugiere que los modelos aprenden representaciones más precisas de la dinámica de la tasa de cambio con más información. Sin embargo, se observa que la magnitud de la reducción del MAE se vuelve menor al añadir cantidades mayores de datos. Adicional, se evidencia que los errores de la predicción de largo plazo son mayores a los de la predicción de corto plazo, lo que puede deberse a la mayor incertidumbre con respecto a un horizonte temporal más amplio.

Se aprecian diferencias significativas en los valores de MAE entre los distintos tipos de modelos para una misma cantidad de datos. Por ejemplo, para cantidades de datos más grandes (12600 y 25200), el modelo ANN tiende a mostrar un MAE en el conjunto de validación relativamente bajo en comparación con otros modelos, lo que sugiere una mayor precisión en sus predicciones sobre datos no vistos. Por otro lado, el modelo LSTM a menudo presenta una mayor variación porcentual del MAE entre el entrenamiento y los conjuntos de prueba/validación

Valores positivos altos en las columnas de variación indican un sobreajuste potencial, donde el modelo se ajusta muy bien a los datos de entrenamiento, pero pierde precisión con datos nuevos. Por ejemplo, algunas pruebas con LSTM muestran aumentos notables en el MAE al pasar del entrenamiento a la prueba, lo cual se evidencia en un aumento del 500 % en la modelación con 3150 datos.

12.1.7. ANÁLISIS DEL COMPORTAMIENTO DEL R2 PARA LARGO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Test	Variación Test - Validación
ann	1575	Serie de tiempo	-0,217853	0,778412	0,956370	457,31%	22,86%
ann	3150	Serie de tiempo	0,959224	0,950091	0,977296	0,95%	2,86%
ann	6300	Serie de tiempo	0,921346	0,971457	0,988842	5,44%	1,79%
ann	12600	Serie de tiempo	0,987036	0,978295	0,994639	0,89%	1,67%
ann	25200	Serie de tiempo	0,995931	0,978011	0,994332	1,80%	1,67%
cnn	1575	Serie de tiempo	0,585033	0,898424	0,885106	53,57%	1,48%
cnn	3150	Serie de tiempo	0,770423	0,938632	0,970841	21,83%	3,43%
cnn	6300	Serie de tiempo	0,843330	0,958767	0,947956	13,69%	1,13%
cnn	12600	Serie de tiempo	0,986343	0,976385	0,992032	1,01%	1,60%
cnn	25200	Serie de tiempo	0,995643	0,979913	0,993953	1,58%	1,43%
lstm	1575	Serie de tiempo	0,681174	0,580248	0,883145	14,82%	52,20%
lstm	3150	Serie de tiempo	0,897034	0,010068	0,844915	98,88%	8292,08%
lstm	6300	Serie de tiempo	0,904638	0,539265	0,927370	40,39%	71,97%
lstm	12600	Serie de tiempo	0,985989	0,965609	0,988130	2,07%	2,33%
lstm	25200	Serie de tiempo	0,997303	0,970627	0,991603	2,67%	2,16%
transformer	1575	Serie de tiempo	-0,222225	0,853775	0,972785	484,19%	13,94%
transformer	3150	Serie de tiempo	0,446274	0,834376	0,963394	86,96%	15,46%
transformer	6300	Serie de tiempo	0,885016	0,937886	0,945514	5,97%	0,81%
transformer	12600	Serie de tiempo	0,983649	0,972799	0,989336	1,10%	1,70%
transformer	25200	Serie de tiempo	0,991177	0,971130	0,990774	2,02%	2,02%

Tabla 7 - Resultados R2 de experimentos de variación de número de registros de entrenamiento para largo plazo

La Tabla 7 describe cómo en el experimento de largo plazo, que a medida que se incrementa la cantidad de datos de entrenamiento, el R2 en los conjuntos de prueba y validación tiende a aumentar, acercándose a 1. También es llamativo ver que algunos resultados son inferiores a 0.50 en coeficiente de correlación, luego, al incrementar la cantidad de datos se puede evidenciar una mejora en el desempeño de los modelos.

CNN y ANN tienden a alcanzar un R2 en el conjunto de validación más alto en comparación con los demás modelos, lo que sugiere una mejor capacidad para modelar la relación entre las variables. Sin embargo, también se logra identificar una variación importante de train a test en esta métrica. En este caso puede deberse a la falta de información o bajo ajuste del modelo en etapas tempranas con una baja cantidad de datos.

La variación de LSTM en los 3150 datos es significativa y toma un comportamiento extraño, dado que se logra evidenciar que para el entrenamiento logra un R2 de casi 90 %, sin embargo, al hacer el test, el R2 llega a ser de sólo el 1 %. Esto puede indicar un evento dentro de la serie de tiempo que no está correctamente especificado y que hace que las características seleccionadas no se acomoden lo suficiente a la serie de tiempo seleccionada.

12.1.8. EVOLUCIÓN DEL MAE AL AUMENTAR LA CANTIDAD DE REGISTROS EN LARGO PLAZO

La Figura 12 muestra la variación del resultado del MAE con respecto a la cantidad de datos con los que cada uno de los modelos (ANN, CNN, LSTM y Transformer) se va entrenando. Los resultados de la visualización corresponden al MAE de validación con datos diarios

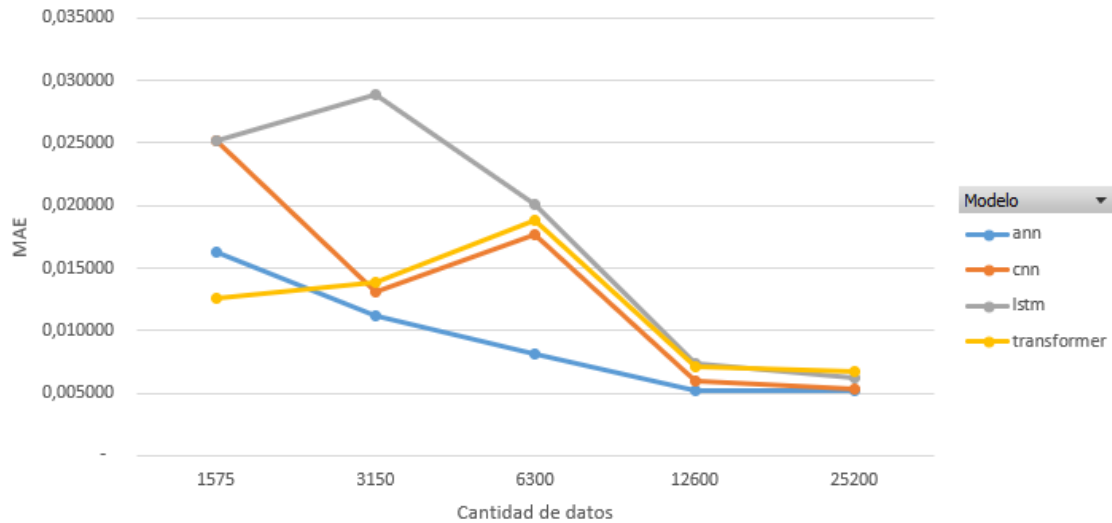


Figura 12 - Desempeño de MAE al aumentar la cantidad de datos en el horizonte de largo plazo

Como lo muestra la Figura 12 todos los modelos tienen una relación inversamente proporcional con respecto a la cantidad de datos usados para el entrenamiento y el MAE de validación para los mismos. Todos los modelos llegan a un rendimiento estándar al alcanzar los 25.200 datos. El modelo con mayor beneficio por la cantidad de datos en el entrenamiento diario es el LSTM, el cual logra acercarse mucho al resto de modelos, a pesar de su rendimiento en la primera iteración con pocos datos.

12.1.9. PREDICCIÓN LOGRADA EN LOS MODELOS ENTRENADOS PARA LARGO PLAZO

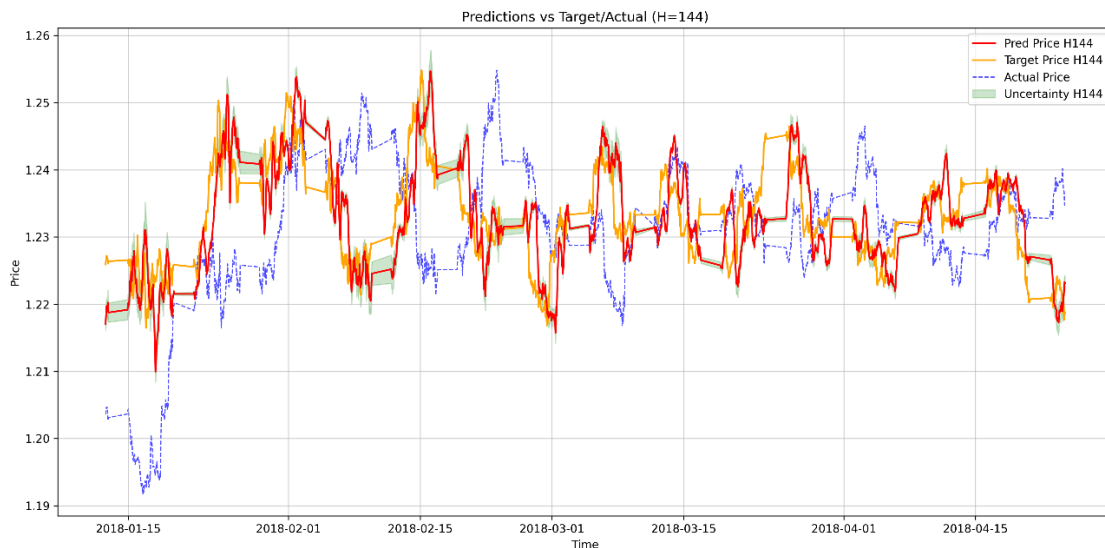


Figura 13 - Predicción para ANN con 25.200 steps con periodos de largo plazo

El modelo ANN, entrenado con 25.200 pasos para predicciones de largo plazo, demostró ser el más efectivo en esta fase del análisis. Tal como se ilustra en la Figura 13, la trayectoria temporal predicha por el modelo muestra correspondencia con el comportamiento observado en los datos. En las métricas de evaluación obtenidas se puede mostrar un MAE reducido y un R2 elevado lo que puede llegar a indicar una buena aproximación del modelo a la serie real.

12.2. EXPERIMENTO DE VARIACIÓN DE NÚMERO Y TIPO DE CARACTERÍSTICAS DE LOS DATOS DE ENTRENAMIENTO

En este experimento se evalúa cómo la adición de diferentes tipos de características (dimensionalidad) a los datos de entrada influye en la precisión y robustez de los modelos predictivos (ANN, CNN, LSTM y Transformers) para la tasa de cambio EUR/USD. A diferencia del experimento anterior (Sección 12), aquí se mantiene constante la cantidad de datos de entrenamiento, utilizando **25,200 registros** del conjunto de entrenamiento disponible (descrito más abajo), cantidad que demostró un buen desempeño en validación en el experimento previo. El objetivo es determinar qué tipo de información adicional aporta más valor a las predicciones de corto plazo (horizonte de 6 horas - corto plazo) y largo plazo (horizonte de 6 días - largo plazo).

Es importante aclarar que la métrica SNR (Signal-to-Noise Ratio) reportada en esta sección se calcula de una manera específica. Los modelos utilizados incorporan una capa final Bayesiana que, para cada punto de predicción, genera múltiples realizaciones (en este caso, 20) con ligeras variaciones. La predicción final reportada es el promedio de estas 20 realizaciones, mientras que la desviación estándar de las mismas se considera una medida de la incertidumbre del modelo en esa predicción. El SNR aquí se calcula como la media de las predicciones producidas por la capa

bayesiana dividida por su desviación estándar (incertidumbre). Por lo tanto, un SNR más alto indica una predicción más "segura" o con menor incertidumbre relativa a su magnitud.

12.2.1. METODOLOGÍAS DE DESCOMPOSICIÓN DE SERIES DE TIEMPO

Antes de alimentar los modelos, y para todos los sub-experimentos de esta sección, la columna close (precio de cierre) de la tasa de cambio EUR/USD fue enriquecida mediante técnicas de descomposición de series de tiempo. Específicamente, se aplicaron las siguientes metodologías para extraer componentes de tendencia, estacionalidad y residuales:

1. **Descomposición STL (Seasonal and Trend decomposition using Loess):** Separa la serie en componentes de tendencia, estacionalidad y residuo utilizando regresión local (Loess). Esta técnica utiliza una ventana de datos pasados para sus cálculos. Se usó una ventana STL de 24 horas pasadas y los valores por defecto para los demás parámetros de la librería statsmodels.
2. **Transformada Wavelet (Db4):** Descompone la señal en diferentes niveles de frecuencia utilizando la wavelet Daubechies 4 (Db4). Esta transformada, al calcular un valor basado en la convolución con la wavelet (que tiene una longitud de 8 ticks en este caso), puede incorporar información del futuro cercano si no se maneja adecuadamente. Para evitar la filtración de datos futuros (*data leakage*), se aplicó un desplazamiento temporal a la derecha de los coeficientes calculados, introduciendo un retraso de 4 ticks respecto a la señal original, pero asegurando que cada valor resultante dependiera únicamente de una ventana de 8 ticks pasados en cada punto. Se usaron los valores por defecto de la librería pywt.
3. **Análisis Multitaper:** Estima el espectro de potencia de la señal utilizando múltiples ventanas o "tapers", lo que permite una mejor resolución espectral y reducción de la varianza en comparación con métodos de una sola ventana. Al igual que STL, utiliza una ventana de datos pasados (de tamaño 144 en este caso) para multiplicar la señal de la ventana por 4 *Sepian Taper Sequences*, que son señales ondulatorias de periodo constante, mutuamente ortogonales entre sí, a las que se calcula el espectro en cada punto usando FFT para obtener la potencia en cada una de las 4 bandas de frecuencia usadas.

Estas tres descomposiciones generaron un total de 11 características adicionales derivadas exclusivamente de la columna close. Las características originales distintas al close (OHLC y variaciones derivadas) y las características adicionales específicas de cada sub-experimento (indicadores técnicos, datos estacionales, etc.) no fueron descompuestas, solo normalizadas junto con las demás.

A continuación, se detallan los resultados para cada grupo de características añadidas. La división general de los datos disponibles fue la siguiente:

- Dataset Total Disponible (EUR/USD Horario):
 - Entrenamiento: 55,318 registros (2005-05-09 a 2014-03-27)
 - Validación: 18,439 registros (2014-03-27 a 2017-03-22)
 - Test: 18,441 registros (2017-03-22 a 2020-03-17)

12.2.2. ADICIÓN DE INDICADORES TÉCNICOS

En este sub-experimento, al dataset base (que incluye OHLC, 8 variaciones derivadas y las 11 características de descomposición) se le añadieron 12 indicadores técnicos calculados sobre la serie de precios EUR/USD horaria. Estos indicadores incluyen señales como MACD, RSI, ADX, Bandas de Bollinger, Oscilador Estocástico, entre otros (excluyendo fecha y OHLC básicos de la lista proporcionada), seleccionados por su bajo Coeficiente de Variación ($CV < 1$) para priorizar señales con mayor relación señal/ruido.

12.2.2.1. DESCRIPCIÓN DE LOS DATOS

- Dataset de Entrenamiento (Utilizado):
 - Filas: 25,200 (seleccionadas del total disponible)
 - Columnas: 31 (8 base + 11 decomp + 12 técnicos)
 - Periodo: Subconjunto de 2005-05-09 a 2009-05-26
- Dataset de Validación:
 - Filas: 18,439
 - Columnas: 38
 - Periodo: 2014-03-27 a 2017-03-22
- Dataset de Test:
 - Filas: 18,441
 - Columnas: 38
 - Periodo: 2017-03-22 a 2020-03-17

12.2.2.2. DESCRIPCIÓN DE LOS MODELOS

Se utilizaron las arquitecturas ANN, CNN, LSTM y Transformer descritas en detalle en la Sección 11 (Figuras 5, 6, 7 y 8). Los modelos fueron entrenados con el dataset descrito anteriormente para predecir los próximos 6 ticks horarios (corto plazo) y los próximos 6 ticks diarios (largo plazo).

12.2.2.3. RESULTADOS: INDICADORES TÉCNICOS

Métricas de Desempeño (MAE) - Adición de Indicadores Técnicos

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Indicadores técnicos	ann	1h	0,000153	0,000143	0,000155	1,31%	8,39%
Indicadores técnicos	ann	1d	0,005445	0,004285	0,004679	-14,07%	9,19%
Indicadores técnicos	cnn	1h	0,000111	0,000087	0,000093	-16,22%	6,90%
Indicadores técnicos	cnn	1d	0,005191	0,005661	0,005794	11,62%	2,35%
Indicadores técnicos	lstm	1h	0,001325	0,000868	0,000952	-28,15%	9,68%
Indicadores técnicos	lstm	1d	0,004802	0,006058	0,006053	26,05%	-0,08%
Indicadores técnicos	transformer	1h	0,001159	0,000806	0,000837	-27,78%	3,85%
Indicadores técnicos	transformer	1d	0,005699	0,005640	0,006191	8,63%	9,77%

Tabla 8 - Resultados MAE de experimentos adición de indicadores técnicos

La Tabla 8 muestra el Error Absoluto Medio (MAE) al añadir indicadores técnicos. En el horizonte corto plazo, el modelo CNN presenta el menor MAE en validación (0.000093), seguido de cerca por ANN (0.000155), indicando una alta precisión en predicciones a corto plazo para estos modelos. LSTM y Transformer muestran errores significativamente mayores en este horizonte. Las variaciones entre Test y Validación son mínimas para todos los modelos en corto plazo, sugiriendo buena generalización en este aspecto.

Para el horizonte largo plazo, el modelo ANN obtiene el MAE más bajo en validación (0.004679), superando a CNN (0.005794), LSTM (0.006053) y Transformer (0.006191). El error es considerablemente mayor que en corto plazo para todos los modelos, lo cual es esperado en predicciones a más largo plazo.

Métricas de Desempeño (R^2) - Adición de Indicadores Técnicos

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Indicadores técnicos	ann	1h	0,999994	0,999979	0,999990	0,00%	0,00%
Indicadores técnicos	ann	1d	0,995899	0,983759	0,994804	-0,11%	1,12%
Indicadores técnicos	cnn	1h	0,999997	0,999993	0,999998	0,00%	0,00%
Indicadores técnicos	cnn	1d	0,996313	0,968068	0,991156	-0,52%	2,38%
Indicadores técnicos	lstm	1h	0,999675	0,999321	0,999767	0,01%	0,04%
Indicadores técnicos	lstm	1d	0,997084	0,966798	0,991375	-0,57%	2,54%
Indicadores técnicos	transformer	1h	0,999727	0,999385	0,999801	0,01%	0,04%
Indicadores técnicos	transformer	1d	0,994593	0,970848	0,991004	-0,36%	2,08%

Tabla 9 - Resultados R^2 de experimentos adición de indicadores técnicos

La Tabla 9 presenta el Coeficiente de Determinación (R^2). Valores cercanos a 1 indican un mejor ajuste del modelo a la varianza de los datos. Para el horizonte corto plazo, todos los modelos alcanzan valores de R^2 extremadamente altos (superiores a 0.999), tanto en entrenamiento como en test y validación. Esto sugiere un ajuste muy bueno a los datos en el corto plazo, lo cual, aunque deseable, podría requerir una revisión para descartar posibles sobre ajustes no evidentes en el MAE o particularidades de la métrica en esta escala. El modelo CNN muestra el R^2 más alto en validación (0.999998).

En el horizonte largo plazo, los valores de R^2 son también altos (generalmente >0.99 en validación, excepto para Transformer), pero muestran una mayor caída entre entrenamiento y test/validación, especialmente para CNN y LSTM (Variación Train-Test > 0.028). El modelo ANN, que tuvo el menor MAE en largo plazo, también presenta el R^2 más alto en validación (0.994804) y una menor caída de rendimiento entre conjuntos, reforzando su buen desempeño a largo plazo con estos indicadores.

Métricas de Desempeño (SNR) - Adición de Indicadores Técnicos

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Indicadores técnicos	ann	1h	3.976,1	4.967,7	5.174,9	30,15%	4,17%
Indicadores técnicos	ann	1d	944,5	1.338,8	1.203,0	27,36%	-10,14%
Indicadores técnicos	cnn	1h	3.718,0	4.912,6	5.059,9	36,09%	3,00%
Indicadores técnicos	cnn	1d	827,7	1.105,1	1.053,8	27,32%	-4,64%
Indicadores técnicos	lstm	1h	2.247,4	2.940,7	3.019,5	34,35%	2,68%
Indicadores técnicos	lstm	1d	553,5	658,6	691,9	25,01%	5,05%
Indicadores técnicos	transformer	1h	1.176,0	1.612,8	1.535,0	30,53%	-4,82%
Indicadores técnicos	transformer	1d	476,8	620,2	584,7	22,62%	-5,72%

Tabla 10 - Resultados SNR de experimentos adición de indicadores técnicos

La Tabla 10 muestra la Relación Señal/Ruido (SNR) de la predicción, donde valores más altos indican menor incertidumbre relativa. En el horizonte corto plazo, los modelos ANN (5174.93) y CNN (5059.95) alcanzan los valores de SNR más altos en validación, sugiriendo predicciones con menor incertidumbre a corto plazo al usar indicadores técnicos. LSTM y Transformer presentan valores considerablemente menores (aproximadamente 40-70% más bajos). Para largo plazo, los valores de SNR disminuyen drásticamente para todos los modelos, indicando un aumento significativo de la incertidumbre en las predicciones a largo plazo, como era de esperar. El SNR de largo plazo para ANN es aproximadamente un 77% menor que su SNR en corto plazo (1202.95 vs 5174.93), y caídas similares se observan en los otros modelos (CNN: ~80%, LSTM: ~77%, Transformer: ~62%). ANN mantiene el SNR más alto en validación para largo plazo (1202.95), alineándose con sus buenos resultados en MAE y R^2 en este horizonte. La consistencia del SNR entre Test y Validación es generalmente buena, con algunas variaciones negativas que indican una ligera reducción de la incertidumbre en el conjunto final.

12.2.2.4. MEJOR PREDICCIÓN: INDICADORES TÉCNICOS

Se pudo observar en los resultados, que el modelo con mejor desempeño para predicciones de corto plazo fue CNN y para largo plazo fue ANN.

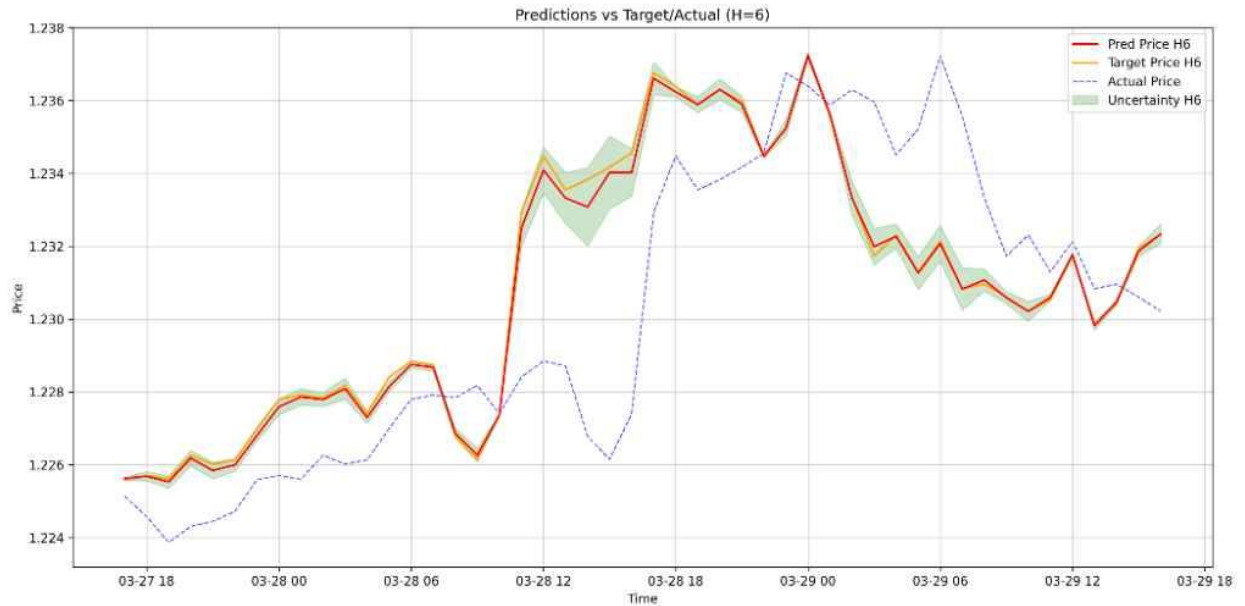


Figura 14 - Comparación entre valores reales y predichos por el modelo CNN utilizando indicadores técnicos - corto plazo

En la Figura 14 se puede observar que la predicción está muy cerca del valor deseado (Target) y predice correctamente el valor real (Actual Price)

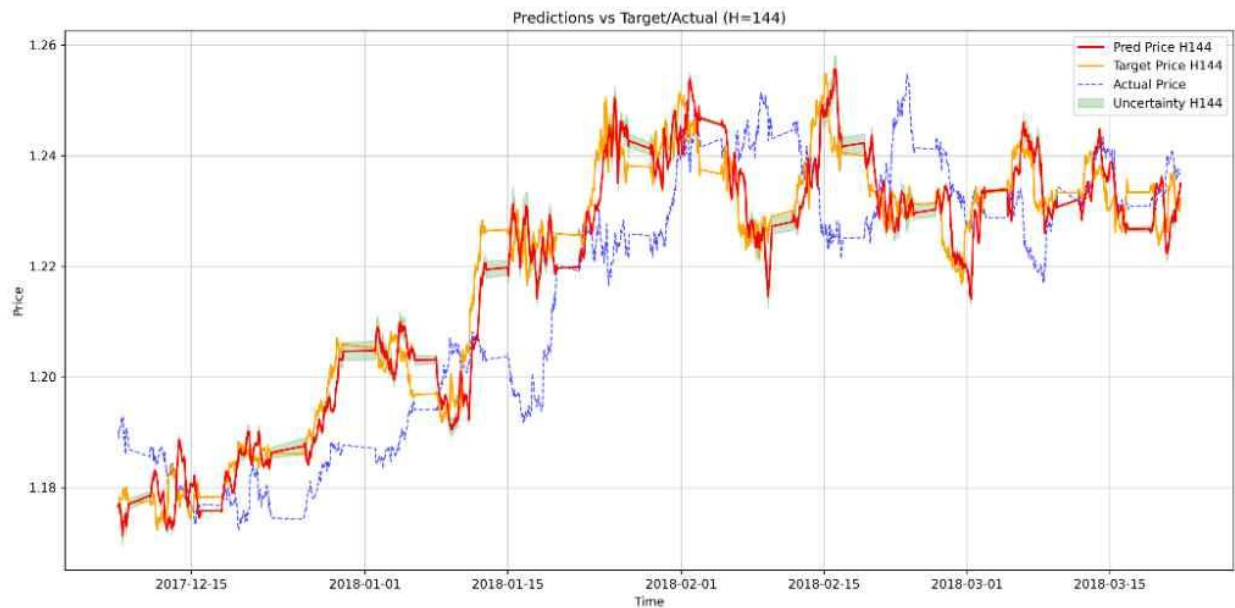


Figura 15 - Comparación entre valores reales y predichos por el modelo ANN utilizando indicadores técnicos – largo plazo

En la Figura 15 se puede observar que, aunque tiene una mayor cantidad de ruido y un mayor error que la predicción de corto plazo (6h, Figura 14), la predicción tiende a seguir el valor Target que se da en anticipación al valor Actual como es el comportamiento deseado.

12.2.3. ADICIÓN DE DATOS ESTACIONALES

En este caso, se añadieron 3 características explícitas para capturar patrones temporales: hora del día (0-23), día de la semana (0-6) y día del mes (1-31).

12.2.3.1. DESCRIPCIÓN DE LOS DATOS

- Dataset de Entrenamiento (Utilizado):
 - Filas: 25,200 (seleccionadas del total disponible: 55,318)
 - Columnas: 22 (8 base + 11 decomp + 3 estacionales)
 - Periodo: Subconjunto de 2005-05-09 a 2009-05-26
- Dataset de Validación:
 - Filas: 18,439
 - Columnas: 22
 - Periodo: 2014-03-27 a 2017-03-22
- Dataset de Test:
 - Filas: 18,441
 - Columnas: 22
 - Periodo: 2017-03-22 a 2020-03-17

12.2.3.2. DESCRIPCIÓN DE LOS MODELOS

Se utilizaron las mismas arquitecturas ANN, CNN, LSTM y Transformer (Sección 11).

12.2.3.3. RESULTADOS: DATOS ESTACIONALES

Métricas de Desempeño (MAE) - Adición de Datos Estacionales

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Dato estacional	ann	1h	0,000140	0,000154	0,000149	6,43%	-3,25%
Dato estacional	ann	1d	0,005530	0,004373	0,004840	-12,48%	10,68%
Dato estacional	cnn	1h	0,000116	0,000086	0,000082	-29,31%	-4,65%
Dato estacional	cnn	1d	0,005432	0,005428	0,005176	-4,71%	-4,64%
Dato estacional	lstm	1h	0,001236	0,000811	0,000883	-28,56%	8,88%
Dato estacional	lstm	1d	0,004757	0,006491	0,006233	31,03%	-3,97%
Dato estacional	transformer	1h	0,001068	0,000690	0,000759	-28,93%	10,00%
Dato estacional	transformer	1d	0,008738	0,006169	0,007370	-15,66%	19,47%

Tabla 11 - Resultados MAE de experimentos adición de datos estacionales

La Tabla 11 evalúa el MAE al añadir datos estacionales. Para corto plazo, CNN vuelve a destacar con el MAE de validación más bajo (0.000082), mostrando una precisión muy alta. ANN le sigue

(0.000149), mientras LSTM y Transformer presentan errores mayores. La generalización (Variación Test-Validación) es excelente para CNN y ANN.

En largo plazo, ANN (0.004840) y CNN (0.005176) obtienen los menores errores de validación, siendo significativamente mejores que LSTM (0.006233) y especialmente Transformer (0.007370), que parece tener dificultades con este horizonte y tipo de datos. Las variaciones negativas Test-Validación para CNN y LSTM sugieren una ligera mejora en el último conjunto de datos no vistos.

Métricas de Desempeño (R^2) - Adición de Datos Estacionales

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Dato estacional	ann	1h	0,999995	0,999975	0,999989	0,00%	0,00%
Dato estacional	ann	1d	0,995837	0,983090	0,994426	-0,14%	1,15%
Dato estacional	cnn	1h	0,999997	0,999994	0,999998	0,00%	0,00%
Dato estacional	cnn	1d	0,995816	0,970955	0,992378	-0,35%	2,21%
Dato estacional	lstm	1h	0,999739	0,999454	0,999812	0,01%	0,04%
Dato estacional	lstm	1d	0,997140	0,959630	0,989437	-0,77%	3,11%
Dato estacional	transformer	1h	0,999762	0,999514	0,999831	0,01%	0,03%
Dato estacional	transformer	1d	0,987053	0,965387	0,986933	-0,01%	2,23%

Tabla 12 - Resultados R^2 de experimentos adición de datos estacionales

La Tabla 12 muestra el R^2 al añadir datos estacionales. Similar al caso con indicadores técnicos, los valores de R^2 para corto plazo son extremadamente altos (≥ 0.9997) para todos los modelos en todos los conjuntos, con CNN alcanzando el valor más alto en validación (0.999998).

Para largo plazo, los R^2 de validación son altos para ANN (0.994426) y CNN (0.992378), pero notablemente inferiores para LSTM (0.989437) y Transformer (0.986933). LSTM muestra la mayor caída entre entrenamiento y test (0.03751). ANN mantiene el R^2 más alto en validación, alineándose con su buen desempeño en MAE para largo plazo.

Métricas de Desempeño (SNR) - Adición de Datos Estacionales

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Dato estacional	ann	1h	3.833,6	4.929,5	4.859,4	26,76%	-1,42%
Dato estacional	ann	1d	901,7	1.284,3	1.125,4	24,81%	-12,37%
Dato estacional	cnn	1h	3.561,2	4.760,8	4.741,9	33,16%	-0,40%
Dato estacional	cnn	1d	777,6	1.027,1	943,7	21,36%	-8,11%
Dato estacional	lstm	1h	2.187,0	2.747,7	2.881,0	31,73%	4,85%
Dato estacional	lstm	1d	516,4	595,3	640,7	24,08%	7,63%
Dato estacional	transformer	1h	1.249,2	1.585,6	1.684,0	34,81%	6,21%
Dato estacional	transformer	1d	255,6	367,3	370,0	44,78%	0,73%

Tabla 13 - Resultados SNR de experimentos adición de datos estacionales

La Tabla 13 presenta el SNR con datos estacionales. En corto plazo, ANN (4859.38) y CNN (4741.91) nuevamente dominan en validación, mostrando la captura de una señal fuerte. LSTM y Transformer quedan rezagados.

Para largo plazo, se observa una caída drástica del SNR en todos los modelos (ANN: ~77%, CNN: ~80%, LSTM: ~78%, Transformer: ~78%), reflejando la mayor incertidumbre a largo plazo. ANN (1125.44) mantiene el SNR más alto en validación para largo plazo, seguido por CNN (943.75). Transformer muestra el SNR más bajo (370.02), lo que concuerda con su bajo desempeño en MAE

y R^2 para este horizonte y tipo de datos. La consistencia entre Test y Validación es buena, con variaciones negativas para ANN y CNN en largo plazo.

12.2.3.4. MEJOR PREDICCIÓN: DATOS ESTACIONALES

Se pudo observar en los resultados de esta sección, que el modelo con mejor desempeño para predicciones de corto plazo fue CNN y para largo plazo fue ANN.

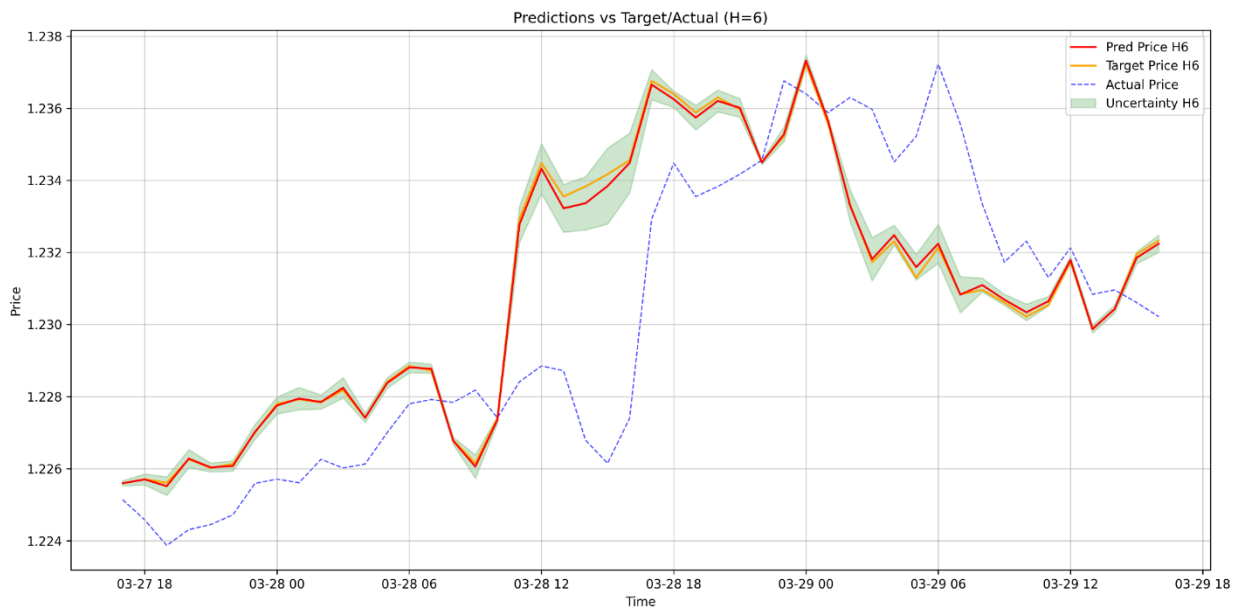


Figura 16 - Comparación entre valores reales y predichos por el modelo CNN adicionando datos estacionales – corto plazo

En la Figura 16, se puede observar un buen ajuste a la señal Target con baja incertidumbre producida por el modelo predictivo CNN en el corto plazo (6h).

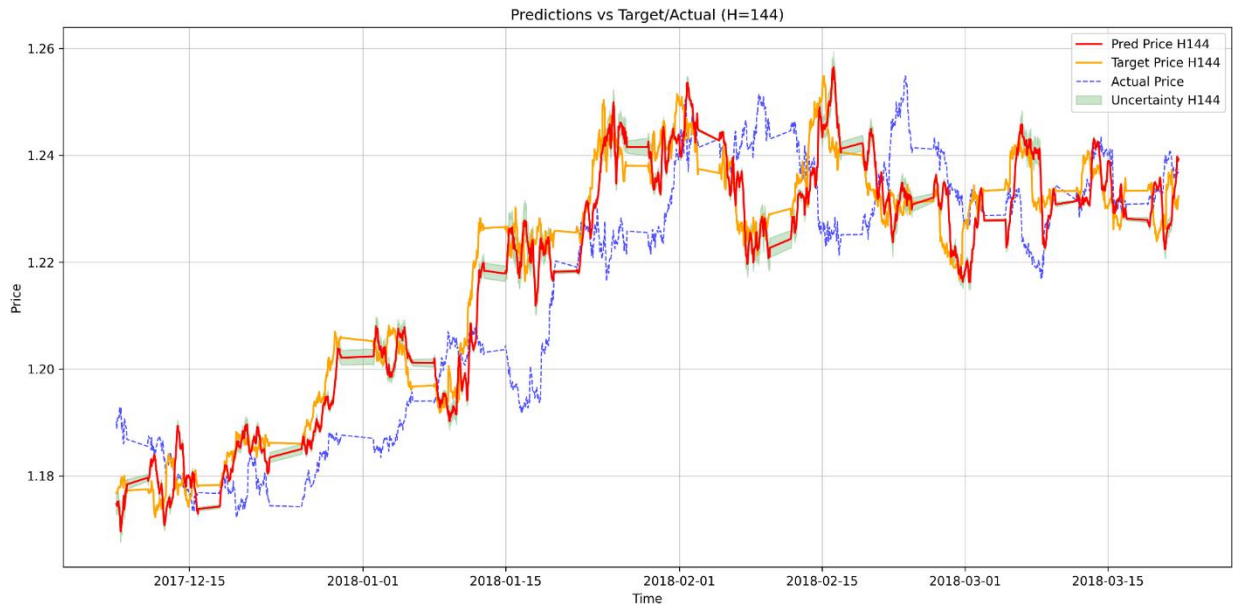


Figura 17 - Comparación entre valores reales y predichos por el modelo ANN adicionando datos estacionales – largo plazo

En la Figura 17, se puede apreciar que el modelo predictivo sigue satisfactoriamente a la señal Target, aunque con cierto nivel de error que se manifiesta como ruido en la predicción.

12.2.4. ADICIÓN DE DATOS DE ALTA FRECUENCIA

Se incorporaron características derivadas de periodicidades más cortas para capturar micro-movimientos recientes. Específicamente, se añadieron los últimos 8 valores de cierre (close) de periodicidad 15 minutos y los últimos 8 valores de cierre de periodicidad 30 minutos, relativos a la hora actual. Esto suma un total de 16 características adicionales.

12.2.4.1. DESCRIPCIÓN DE LOS DATOS

- Dataset de Entrenamiento (Utilizado):
 - Filas: 25,200 (seleccionadas del total disponible: 55,318)
 - Columnas: 35 (8 base + 11 decomp + 16 alta frec.)
 - Periodo: Subconjunto de 2005-05-09 a 2009-05-26
- Dataset de Validación:
 - Filas: 18,439
 - Columnas: 35
 - Periodo: 2014-03-27 a 2017-03-22
- Dataset de Test:
 - Filas: 18,441
 - Columnas: 35
 - Periodo: 2017-03-22 a 2020-03-17

12.2.4.2. DESCRIPCIÓN DE LOS MODELOS

Se utilizaron las mismas arquitecturas ANN, CNN, LSTM y Transformer (Sección 11).

12.2.4.3. RESULTADOS: DATOS DE ALTA FRECUENCIA

Métricas de Desempeño (MAE) - Adición de Datos de Alta Frecuencia

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Alta frecuencia	ann	1h	0,000154	0,000162	0,000154	0,00%	-4,94%
Alta frecuencia	ann	1d	0,005564	0,004273	0,004745	-14,72%	11,05%
Alta frecuencia	cnn	1h	0,000125	0,000084	0,000101	-19,20%	20,24%
Alta frecuencia	cnn	1d	0,005564	0,005978	0,005334	-4,13%	-10,77%
Alta frecuencia	lstm	1h	0,001417	0,000928	0,001003	-29,22%	8,08%
Alta frecuencia	lstm	1d	0,005216	0,005489	0,005906	13,23%	7,60%
Alta frecuencia	transformer	1h	0,001279	0,000829	0,000869	-32,06%	4,83%
Alta frecuencia	transformer	1d	0,004364	0,005819	0,005674	30,02%	-2,49%

Tabla 14 - Resultados MAE de experimentos adición de datos alta frecuencia

La Tabla 14 presenta el MAE al incorporar datos de alta frecuencia. En corto plazo, CNN (0.000101) vuelve a mostrar el error de validación más bajo, aunque ANN (0.000154) también es competitivo. LSTM y Transformer siguen mostrando errores mayores.

Para largo plazo, ANN (0.004745) lidera claramente en MAE de validación, seguido por CNN (0.005334). Transformer (0.005674) mejora respecto a otros escenarios largo plazo, superando a LSTM (0.005906) en esta ocasión. Las variaciones Test-Validación negativas para CNN y Transformer sugieren buena generalización final.

Métricas de Desempeño (R^2) - Adición de Datos de Alta Frecuencia

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Alta frecuencia	ann	1h	0,999994	0,999976	0,999990	0,00%	0,00%
Alta frecuencia	ann	1d	0,995692	0,983858	0,994669	-0,10%	1,10%
Alta frecuencia	cnn	1h	0,999997	0,999994	0,999998	0,00%	0,00%
Alta frecuencia	cnn	1d	0,995611	0,960190	0,990953	-0,47%	3,20%
Alta frecuencia	lstm	1h	0,999672	0,999317	0,999767	0,01%	0,05%
Alta frecuencia	lstm	1d	0,996523	0,972836	0,991069	-0,55%	1,87%
Alta frecuencia	transformer	1h	0,999560	0,999142	0,999706	0,01%	0,06%
Alta frecuencia	transformer	1d	0,997592	0,965769	0,991485	-0,61%	2,66%

Tabla 15 - Resultados R^2 de experimentos adición de datos alta frecuencia

La Tabla 15 muestra el R^2 con datos de alta frecuencia. Los resultados son muy similares a los obtenidos con indicadores técnicos y datos estacionales. En corto plazo, los R^2 son extremadamente altos para todos los modelos, liderados por CNN (0.999998). En largo plazo, ANN (0.994669) presenta el R^2 de validación más alto y la menor caída Train-Test/Validación entre los modelos con $R^2 > 0.99$. CNN, LSTM y Transformer muestran caídas más significativas entre entrenamiento y los experimentos posteriores, especialmente CNN (0.03542 Train-Test), sugiriendo mayor sobreajuste potencial.

Métricas de Desempeño (SNR) - Adición de Datos de Alta Frecuencia

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Alta frecuencia	ann	1h	3.788,7	4.734,5	5.028,2	32,72%	6,21%
Alta frecuencia	ann	1d	976,2	1.388,8	1.281,4	31,26%	-7,73%
Alta frecuencia	cnn	1h	3.646,6	4.840,3	4.949,0	35,72%	2,25%
Alta frecuencia	cnn	1d	769,8	927,0	914,0	18,74%	-1,40%
Alta frecuencia	lstm	1h	2.154,8	2.780,9	2.881,5	33,72%	3,62%
Alta frecuencia	lstm	1d	557,9	711,1	674,4	20,89%	-5,16%
Alta frecuencia	transformer	1h	1.311,1	1.619,5	1.708,9	30,34%	5,52%
Alta frecuencia	transformer	1d	705,4	958,2	1.035,4	46,79%	8,06%

Tabla 16 - Resultados SNR de experimentos adición de datos alta frecuencia

La Tabla 16 evalúa el SNR (inverso de la incertidumbre relativa) al añadir datos de alta frecuencia. En corto plazo, ANN (5028.23) y CNN (4949.04) mantienen los SNR de validación más altos, indicando predicciones de baja incertidumbre. En largo plazo, el SNR cae significativamente para todos los modelos (ANN: ~75%, CNN: ~81%, LSTM: ~77%, Transformer: ~39%), reflejando el aumento de la incertidumbre a largo plazo. ANN (1281.42) conserva el SNR más alto en validación. Notablemente, Transformer (1035.43) muestra la menor caída porcentual de SNR entre corto plazo y largo plazo (~39%), lo que sugiere que los datos de alta frecuencia podrían ayudar a reducir la incertidumbre relativa de este modelo en predicciones largas, aunque su precisión (MAE) no sea la mejor. LSTM (674.38) presenta el SNR más bajo.

12.2.4.4. MEJOR PREDICCIÓN: DATOS DE ALTA FRECUENCIA

Al igual que en con los otros tipos de datos de la sección 12.8, para datos de alta frecuencia, el mejor modelo predictivo de mejor desempeño para predicciones de corto plazo fué CNN y para largo plazo fué ANN.

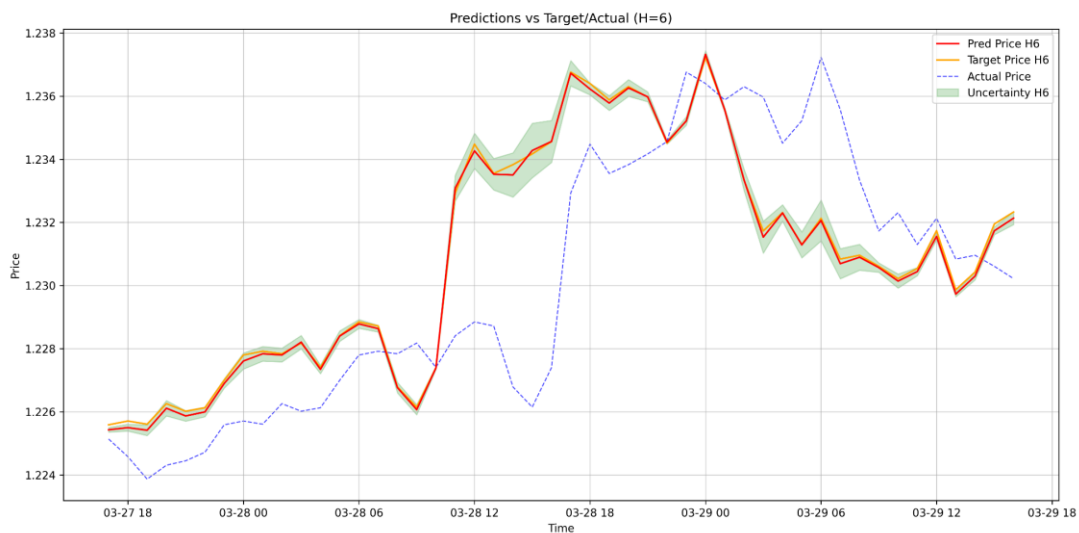


Figura 18 - Comparación entre valores reales y predichos por el modelo CNN adicionando datos de alta frecuencia – corto plazo

En la figura 18 se puede apreciar que el modelo CNN generó predicciones de corto plazo con baja incertidumbre y un ajuste muy bueno a la señal de Target que anticipa el valor furo de la señal Actual Price (Close).

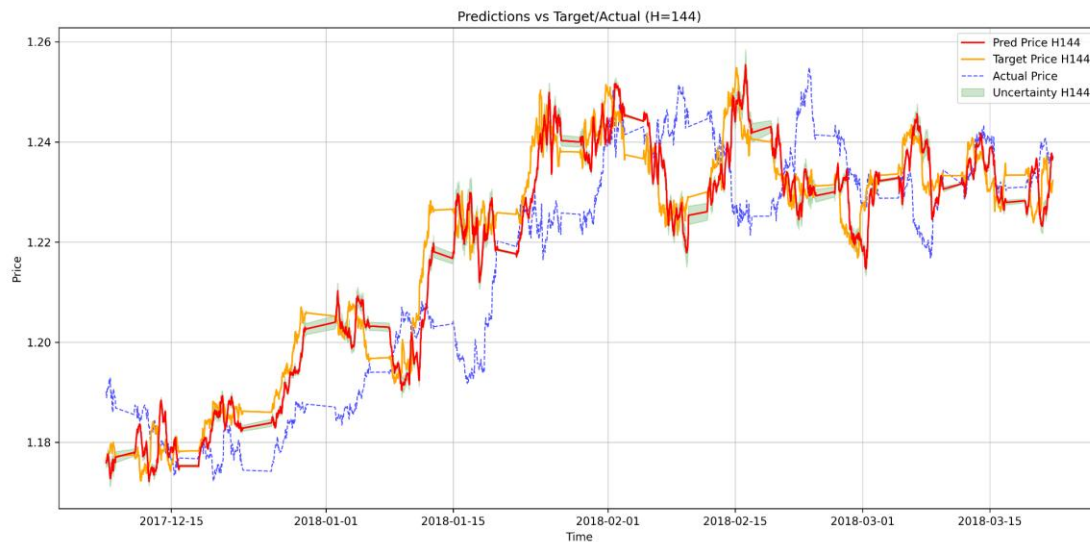


Figura 19 - Comparación entre valores reales y predichos por el modelo ANN adicionando datos de alta frecuencia – largo plazo

Se puede apreciar la menor exactitud en la predicción debido al mayor error y por tanto mayor nivel de ruido que en las predicciones de corto plazo, sin embargo, la predicción se ajusta a la señal de Target aproximadamente.

12.2.5. ADICIÓN DE INDICADORES FUNDAMENTALES

En esta etapa se incorporaron indicadores fundamentales que reflejan el sentimiento del mercado y la volatilidad: el índice S&P 500 y el índice de volatilidad VIX. Dado que estos datos tienen periodicidad diaria, se replicó el valor diario para todas las horas correspondientes a ese día. Se añadieron 2 características.

12.2.5.1. DESCRIPCIÓN DE LOS DATOS

- Dataset de Entrenamiento (Utilizado):
 - Filas: 25,200 (seleccionadas del total disponible: 55,318)
 - Columnas: 21 (8 base + 11 decomp + 2 fundamentales)
 - Periodo: Subconjunto de 2005-05-09 a 2009-05-26
- Dataset de Validación:
 - Filas: 18,439
 - Columnas: 21
 - Periodo: 2014-03-27 a 2017-03-22
- Dataset de Test:
 - Filas: 18,441

- Columnas: 21
- Periodo: 2017-03-22 a 2020-03-17

12.2.5.2. DESCRIPCIÓN DE LOS MODELOS

Se utilizaron las mismas arquitecturas ANN, CNN, LSTM y Transformer (Sección 11).

12.2.5.3. RESULTADOS: INDICADORES FUNDAMENTALES

Métricas de Desempeño (MAE) - Adición de Indicadores Fundamentales

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Fundamentales	ann	1h	0,000184	0,000181	0,000188	2,17%	3,87%
Fundamentales	ann	1d	0,005521	0,004358	0,004791	-13,22%	9,94%
Fundamentales	cnn	1h	0,000145	0,000144	0,000122	-15,86%	-15,28%
Fundamentales	cnn	1d	0,005463	0,005902	0,005668	3,75%	-3,96%
Fundamentales	lstm	1h	0,001079	0,000797	0,000798	-26,04%	0,13%
Fundamentales	lstm	1d	0,004770	0,005580	0,006048	26,79%	8,39%
Fundamentales	transformer	1h	0,001395	0,000994	0,001027	-26,38%	3,32%
Fundamentales	transformer	1d	0,006508	0,005829	0,007039	8,16%	20,76%

Tabla 17 - Resultados MAE de experimentos adición de indicadores fundamentales

La Tabla 17 presenta el MAE al añadir sólo indicadores fundamentales (S&P 500, VIX). En corto plazo, CNN (0.000122) logra el menor error de validación, seguido por ANN (0.000188). LSTM y Transformer muestran errores mayores. La variación Test-Validación negativa para CNN es un buen signo de generalización.

Para largo plazo, ANN (0.004791) es el claro ganador en MAE de validación. CNN (0.005668) es el segundo mejor, mientras que LSTM (0.006048) y Transformer (0.007039) tienen un desempeño inferior. La variación Test-Validación negativa de CNN en largo plazo es también notable.

Métricas de Desempeño (R²) - Adición de Indicadores Fundamentales

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Fundamentales	ann	1h	0,999994	0,999973	0,999989	0,00%	0,00%
Fundamentales	ann	1d	0,995798	0,983123	0,994649	-0,12%	1,17%
Fundamentales	cnn	1h	0,999996	0,999981	0,999996	0,00%	0,00%
Fundamentales	cnn	1d	0,995844	0,970647	0,992012	-0,38%	2,20%
Fundamentales	lstm	1h	0,999802	0,999497	0,999842	0,00%	0,03%
Fundamentales	lstm	1d	0,997114	0,971396	0,990838	-0,63%	2,00%
Fundamentales	transformer	1h	0,999504	0,998934	0,999622	0,01%	0,07%
Fundamentales	transformer	1d	0,992862	0,968717	0,987702	-0,52%	1,96%

Tabla 18 - Resultados R² de experimentos adición de indicadores fundamentales

La Tabla 18 muestra el R² con indicadores fundamentales. En corto plazo, los valores son nuevamente muy altos (>0.9995), con CNN (0.999996) liderando en validación. En largo plazo, ANN (0.994649) obtiene el mejor R² de validación y la menor caída de rendimiento entre conjuntos. CNN, LSTM y Transformer muestran caídas más pronunciadas entre Train y Test/Validación, sugiriendo que la información fundamental por sí sola podría no ser suficiente para que estos

modelos generalicen tan bien a largo plazo como ANN, o que ANN es menos propenso al sobreajuste en este escenario.

Métricas de Desempeño (SNR) - Adición de Indicadores Fundamentales

Adición	Modelo	Típos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Fundamentales	ann	1h	3.709,2	4.633,0	4.773,5	28,69%	3,03%
Fundamentales	ann	1d	968,7	1.401,5	1.237,6	27,76%	-11,70%
Fundamentales	cnn	1h	3.340,7	4.482,2	4.488,7	34,36%	0,15%
Fundamentales	cnn	1d	773,6	427,5	638,0	-17,53%	49,24%
Fundamentales	lstm	1h	2.229,7	2.190,2	2.716,5	21,83%	24,03%
Fundamentales	lstm	1d	554,1	591,2	663,5	19,75%	12,23%
Fundamentales	transformer	1h	1.507,1	1.713,6	1.879,9	24,73%	9,71%
Fundamentales	transformer	1d	497,1	705,9	724,1	45,68%	2,59%

Tabla 19 - Resultados SNR de experimentos adición de indicadores fundamentales

La Tabla 19 presenta el SNR (inverso de la incertidumbre relativa) al añadir indicadores fundamentales. En corto plazo, ANN (4773.50) y CNN (4488.72) muestran los SNR de validación más altos, indicando la menor incertidumbre.

Para largo plazo, el SNR vuelve a caer drásticamente (ANN: ~74%, CNN: ~86%, LSTM: ~75%, Transformer: ~61%), evidenciando el aumento de incertidumbre. ANN (1237.57) mantiene el SNR más alto en validación, seguido por Transformer (724.14). El SNR de CNN en Test (427.48) es llamativamente bajo comparado con Train y Validación, y su variación Train-Test es fuertemente negativa (-346.11), lo que podría indicar inestabilidad o dificultad del modelo CNN para extraer una señal clara y consistente a largo plazo solo con estos datos fundamentales.

12.2.5.4. MEJOR PREDICCIÓN: INDICADORES FUNDAMENTALES

Al igual que en las secciones anteriores para otros tipos de datos, el modelo con mejor desempeño en el corto plazo fue CNN mientras que en las predicciones de largo plazo, el modelo ANN tuvo las mejores predicciones entre los modelos probados, como se aprecia en las Figuras 20 y 21.

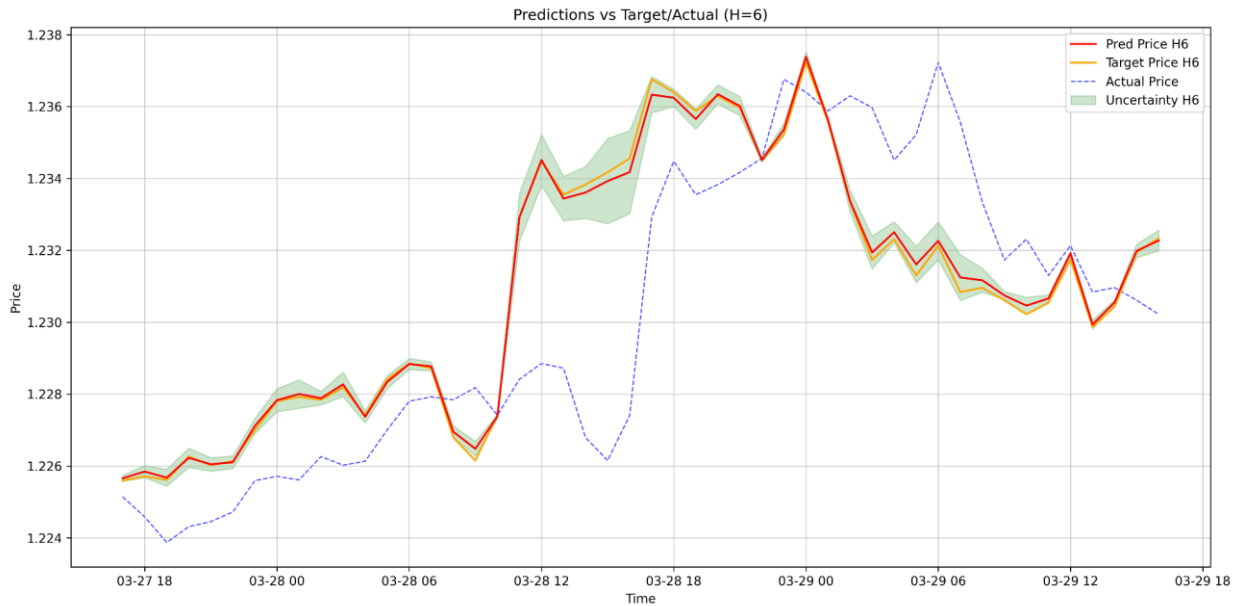


Figura 20 - Comparación entre valores reales y predichos por el modelo CNN adicionando indicadores fundamentales - corto plazo

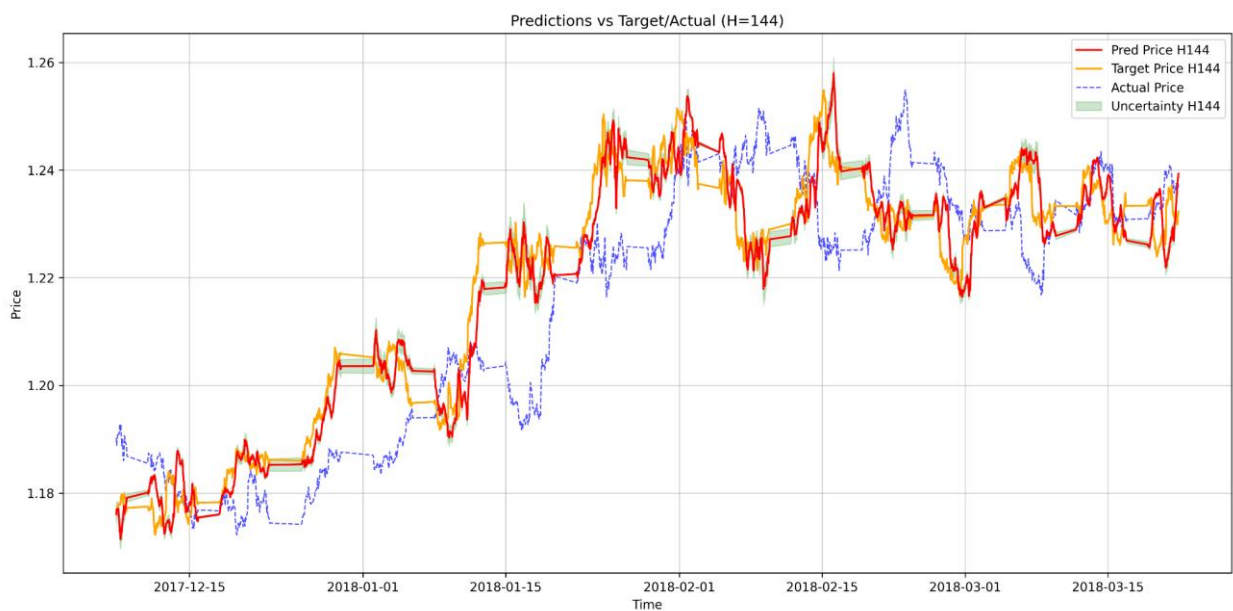


Figura 21 - Comparación entre valores reales y predichos por el modelo ANN adicionando indicadores fundamentales largo plazo Adición de todas las características al modelo

12.2.6. ADICIÓN DE TODAS LAS CARACTERÍSTICAS

Finalmente, en esta etapa se evaluó el desempeño de cada uno de los modelos cuando le fueron adicionados todos los tipos de datos que fueron evaluados de forma independiente en las anteriores 4 subsecciones. Al usar todas las características juntas como entrada, se produjo una

mejora en el desempeño predictivo de todos los modelos respecto al uso de cada característica individualmente.

12.2.6.1. DESCRIPCIÓN DE LOS MODELOS

Se utilizaron las mismas arquitecturas ANN, CNN, LSTM y Transformer (Sección 10) para los extractores de características de cada modelo (primeras dos capas). Pero debido a que la siguiente sección involucra el uso de un extractor de características entrando externamente con un **autoencoder**, se optó por usar cabezales predictivos fijos compuestos por dos capas convolucionales más una capa LSTM y una capa bayesiana de salida, mientras se usan los extractores de características que se describen en la sección 12.10 para cada modelo. De esta manera los resultados de la sección 12.10 se podrán comparar con los resultados de los modelos con extractor entrenado externamente (sección 13), un fragmento del model plot usado, que muestra los cabezales predictivos descritos en la Figura 22 (el diagrama del modelo completo se encuentra en el **Anexo E**).

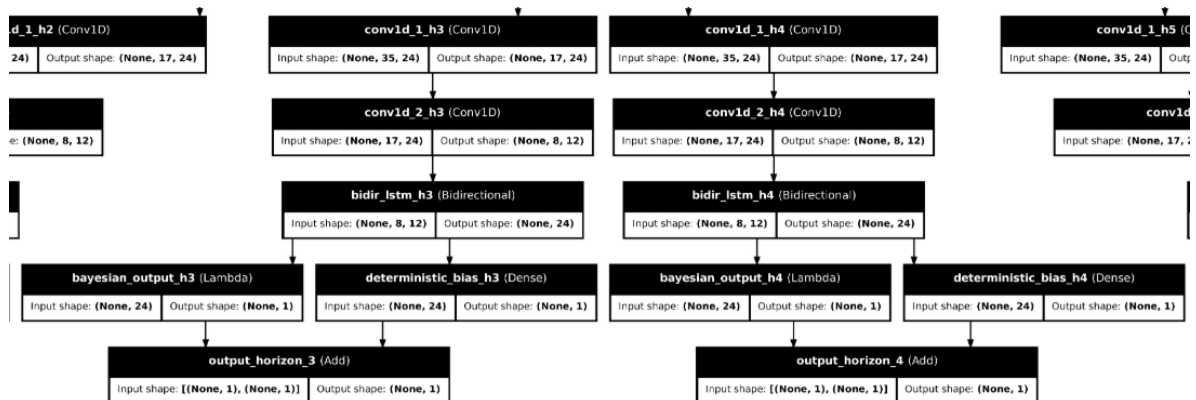


Figura 22 – Cabezales predictivos usados en experimentos de sección 12.10 y 13 (Fragmento de model plot generado con Keras)

12.2.6.2. RESULTADOS: USO DE TODAS LAS CARACTERÍSTICAS

Métricas de Desempeño (MAE) - Adición de todas las características

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Todos	ann	1h	0,000163	0,000110	0,000120	-26,38%	9,09%
Todos	ann	1d	0,004818	0,003000	0,004421	-8,24%	47,37%
Todos	cnn	1h	0,000110	0,000109	0,000126	14,55%	15,60%
Todos	cnn	1d	0,004965	0,003624	0,005405	8,86%	49,14%
Todos	lstm	1h	0,001630	0,001004	0,001380	-15,34%	37,45%
Todos	lstm	1d	0,004737	0,004036	0,005305	11,99%	31,44%
Todos	transformer	1h	0,001342	0,000849	0,001218	-9,24%	43,46%
Todos	transformer	1d	0,007187	0,004199	0,006240	-13,18%	48,61%

Tabla 20 - Resultados MAE de experimentos adición de todas las características

La Tabla 20 presenta el MAE al añadir todas las características. En corto plazo, ANN (0.000120) logra el menor error de validación, seguido por CNN (0.000126). LSTM y Transformer muestran errores

mayores. La variación Test-Validación es la más pequeña para ANN, lo cual puede indicar que es el modelo que mejor generaliza.

Para largo plazo, ANN (0.004791) es el claro ganador en MAE de validación. CNN (0.005668) es el segundo mejor, mientras que LSTM (0.006048) y Transformer (0.007039) tienen un desempeño inferior. La variación Test-Validación negativa de CNN en largo plazo es también notable.

Métricas de Desempeño (R^2) - Adición de Todas las características

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Todos	ann	1h	0,999994	0,999898	0,999967	0,00%	0,01%
Todos	ann	1d	0,996628	0,919071	0,960133	-3,66%	4,47%
Todos	cnn	1h	0,999997	0,999891	0,999960	0,00%	0,01%
Todos	cnn	1d	0,996514	0,885947	0,940690	-5,60%	6,18%
Todos	lstm	1h	0,999566	0,990441	0,995534	-0,40%	0,51%
Todos	lstm	1d	0,996949	0,860508	0,941893	-5,52%	9,46%
Todos	transformer	1h	0,999589	0,990636	0,995539	-0,41%	0,49%
Todos	transformer	1d	0,991879	0,838142	0,914665	-7,78%	9,13%

Tabla 21 - Resultados R^2 de experimentos adición de todas las características

La Tabla 21 muestra el R^2 con todas las características. En corto plazo, los valores son nuevamente muy altos (>0.9995), con CNN (0.999996) liderando en validación. En largo plazo, ANN (0.994649) obtiene el mejor R^2 de validación y la menor caída de rendimiento entre conjuntos.

Métricas de Desempeño (SNR) - Adición de Todas las características

Adición	Modelo	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
Todos	ann	1h	4.338,19	5.827,09	4.752,11	9,54%	-18,45%
Todos	ann	1d	1.123,37	1.631,31	1.258,29	12,01%	-22,87%
Todos	cnn	1h	3.563,83	5.174,38	3.921,74	10,04%	-24,21%
Todos	cnn	1d	822,41	1.162,75	914,58	11,21%	-21,34%
Todos	lstm	1h	2.549,99	3.160,76	2.710,79	6,31%	-14,24%
Todos	lstm	1d	588,09	732,83	672,08	14,28%	-8,29%
Todos	transformer	1h	13.109,36	10.059,81	13.009,56	-0,76%	29,32%
Todos	transformer	1d	462,06	668,93	549,38	18,90%	-17,87%

Tabla 22 - Resultados SNR de experimentos adición de todas las características

La Tabla 22 presenta el SNR al añadir todas las características. En corto plazo, ANN (4773.50) y CNN (4488.72) muestran los valores más altos en validación. Para largo plazo, ANN (1237.57) vuelve a liderar, seguido por Transformer (724.14) y LSTM (663.49).

12.2.6.3. MEJOR PREDICCIÓN: TODAS LAS CARACTERÍSTICAS

Al igual que con las características adicionales por separado, la adición de estas características juntas (seasonal, high frequency, technical and fundamental indicators), el modelo que tuvo mejor desempeño predictivo para el corto plazo fué usando extractores de características (dos primeras capas) con CNN, mientras que para el largo plazo fué usando extractores con ANN.

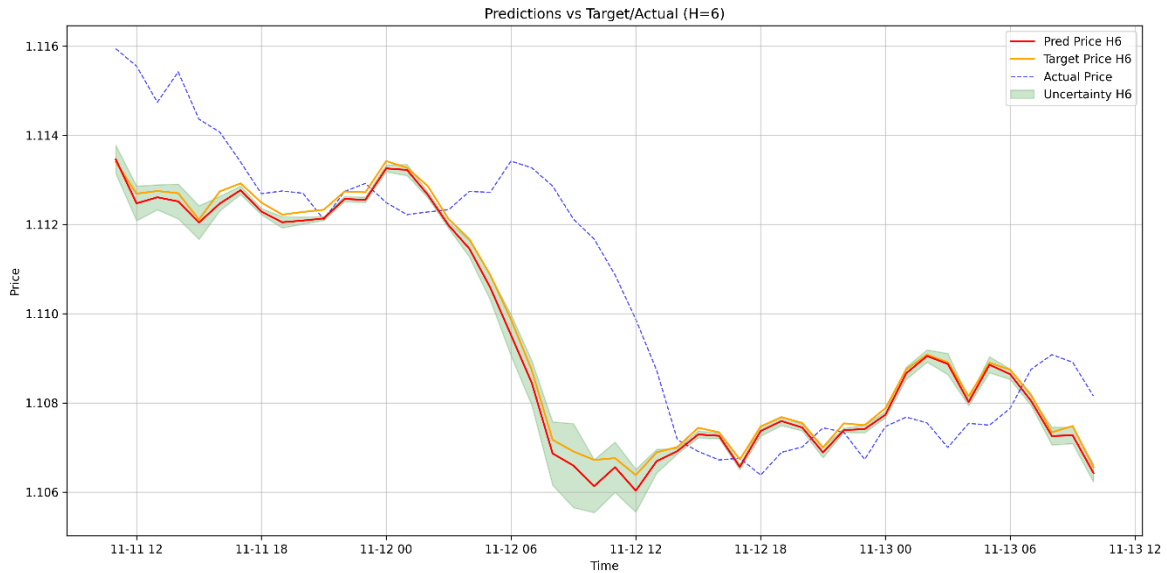


Figura 23 - Comparación entre valores reales y predichos por el modelo ANN adicionando todas las características - corto plazo

En la figura 23 se aprecia una predicción con muy baja incertidumbre que se ajusta muy bien a la señal de Target.

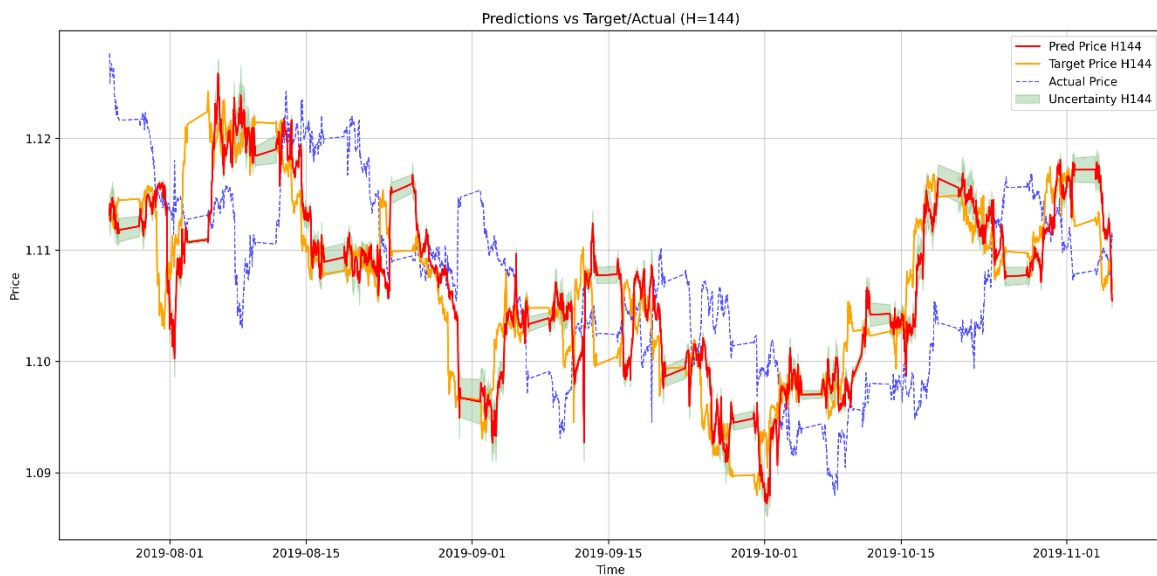


Figura 24 - Comparación entre valores reales y predichos por el modelo ANN adicionando todas las características largo plazo

En la Figura 24 se puede ver que la predicción de largo plazo se ajusta a la señal de Target, pero con mayor nivel de ruido que la predicción de corto plazo.

12.2.7. COMPARACIÓN CON RESULTADOS DEL ANTERIOR EXPERIMENTO (VARIACIÓN DE CANTIDAD DE DATOS)

A continuación, se realiza una comparativa porcentual del mejor MAE de validación obtenido en cada subgrupo de este experimento (variación de características) con respecto al mejor MAE de validación obtenido en el experimento previo (variación de cantidad de datos) utilizando 25,200 registros. La línea base de la Variación de Cantidad de Datos es: MAE Validación corto plazo (CNN) = 0.000134 y MAE Validación largo plazo (ANN) = 0.005175. Un porcentaje negativo indica una mejora (reducción del error) respecto a la línea base.

12.2.7.1. COMPARATIVA VS ADICIÓN INDICADORES TÉCNICOS

Modelo	Horizonte	MAE Validación (Fase 2 - Ind. Técnicos)	MAE Base (Fase 1 - Mejor Modelo*)	Variación MAE (%)	Mejora / Empeora
ANN	h6	0.000155	0.000134	+15.67%	Empeora
ANN	h144	0.004679	0.005175	-9.58%	Mejora
CNN	h6	0.000093	0.000134	-30.60%	Mejora
CNN	h144	0.005794	0.005175	+11.96%	Empeora
LSTM	h6	0.000952	0.000134	+610.45%	Empeora
LSTM	h144	0.006053	0.005175	+16.97%	Empeora
Transformer	h6	0.000837	0.000134	+524.63%	Empeora
Transformer	h144	0.006191	0.005175	+19.63%	Empeora

Tabla 23 - Comparativa MAE Validación - Variación Cantidad de Datos vs Experimento con Indicadores Técnicos

*MAE Base de experimento de Variación de Cantidad de Datos corresponde al mejor MAE de Validación con 25,200 registros: CNN para corto plazo (0.000134) y ANN para largo plazo (0.005175).

- Horizonte corto plazo: El mejor modelo en este experimento fue CNN con MAE Validación = 0.000093. Esto representa una mejora del 30.60% respecto al mejor MAE corto plazo del experimento de Variación de Cantidad de Datos.
- Horizonte largo plazo: El mejor modelo en este experimento fue ANN con MAE Validación = 0.004679. Esto representa una mejora del 9.58% respecto al mejor MAE largo plazo del experimento de Variación de Cantidad de Datos.

La adición de indicadores técnicos mejoró el MAE del mejor modelo en ambos horizontes.

12.2.7.2. COMPARATIVA VS ADICIÓN DE DATOS ESTACIONALES

Modelo	Horizonte	MAE Validación (Fase 2 - Estacionales)	MAE Base (Fase 1 - Mejor Modelo*)	Variación MAE (%)	Mejora/E mpeora
ANN	h6	0.000149	0.000134	+11.19%	Empeora
ANN	h144	0.004840	0.005175	-6.47%	Mejora
CNN	h6	0.000082	0.000134	-38.81%	Mejora
CNN	h144	0.005176	0.005175	+0.02%	Empeora
LSTM	h6	0.000883	0.000134	+558.96%	Empeora
LSTM	h144	0.006233	0.005175	+20.44%	Empeora
Transformer	h6	0.000759	0.000134	+466.42%	Empeora
Transformer	h144	0.007370	0.005175	+42.42%	Empeora

Tabla 24 - Comparativa MAE Validación - Experimento de Variación de Cantidad de Datos vs Experimento con Datos Estacionales

*MAE Base experimento de Variación de Cantidad de Datos corresponde al mejor MAE de Validación con 25,200 registros: CNN para corto plazo (0.000134) y ANN para largo plazo (0.005175).

- Horizonte corto plazo: El mejor modelo en este experimento fue CNN con MAE Validación = 0.000082. Esto representa una mejora del 38.81% respecto al mejor MAE corto plazo del experimento de Variación de Cantidad de Datos.
- Horizonte largo plazo: El mejor modelo en este experimento fue ANN con MAE Validación = 0.004840. Esto representa una mejora del 6.47% respecto al mejor MAE largo plazo del experimento de Variación de Cantidad de Datos.

La adición de datos estacionales también mejoró el MAE del mejor modelo en ambos horizontes, logrando la mayor mejora porcentual para corto plazo.

12.2.7.3. COMPARATIVA VS ADICIÓN DE DATOS DE ALTA FRECUENCIA

Modelo	Horizonte	MAE Validación (Fase 2 - Alta Frec.)	MAE Base (Fase 1 - Mejor Modelo*)	Variación MAE (%)	Mejora / Empeora
ANN	h6	0.000154	0.000134	+14.93%	Empeora
ANN	h144	0.004745	0.005175	-8.31%	Mejora
CNN	h6	0.000101	0.000134	-24.63%	Mejora
CNN	h144	0.005334	0.005175	+3.07%	Empeora
LSTM	h6	0.001003	0.000134	+648.51%	Empeora
LSTM	h144	0.005906	0.005175	+14.12%	Empeora
Transformer	h6	0.000869	0.000134	+548.51%	Empeora
Transformer	h144	0.005674	0.005175	+9.64%	Empeora

Tabla 25 - Comparativa MAE Validación - Variación Cantidad de Datos vs Experimento con adición de datos de Alta Frecuencia

*MAE Base Experimento de Variación de Cantidad de Datos corresponde al mejor MAE de Validación con 25,200 registros: CNN para corto plazo (0.000134) y ANN para largo plazo (0.005175).

- Horizonte corto plazo: El mejor modelo en este experimento fue CNN con MAE Validación = 0.000101. Esto representa una mejora del 24.63% respecto al mejor MAE corto plazo del experimento de Variación de Cantidad de Datos.
- Horizonte largo plazo: El mejor modelo en este experimento fue ANN con MAE Validación = 0.004745. Esto representa una mejora del 8.31% respecto al mejor MAE largo plazo del experimento de Variación de Cantidad de Datos.

La adición de datos de alta frecuencia mejoró el MAE del mejor modelo en ambos horizontes.

12.2.7.4. COMPARATIVA VS ADICIÓN DE INDICADORES FUNDAMENTALES

Modelo	Horizonte	MAE Validación (Fase 2 - Fundamentales)	MAE Base (Fase 1 - Mejor Modelo*)	Variación MAE (%)	Mejora / Empeora
ANN	1h	0.000188	0.000134	+40.30%	Empeora
ANN	1d	0.004791	0.005175	-7.42%	Mejora
CNN	1h	0.000122	0.000134	-8.96%	Mejora
CNN	1d	0.005668	0.005175	+9.53%	Empeora
LSTM	1h	0.000798	0.000134	+495.52%	Empeora
LSTM	1d	0.006048	0.005175	+16.87%	Empeora
Transformer	1h	0.001027	0.000134	+666.42%	Empeora
Transformer	1d	0.007039	0.005175	+36.02%	Empeora

Tabla 26 - Comparativa MAE Validación - Experimento de Variación de Cantidad de Datos. vs Experimento de Adición de Datos Fundamentales

*MAE Base experimento de Variación de Cantidad de Datos. corresponde al mejor MAE de Validación con 25,200 registros: CNN para corto plazo (0.000134) y ANN para largo plazo (0.005175).

- Horizonte corto plazo: El mejor modelo en este experimento fue CNN con MAE Validación = 0.000122. Esto representa una mejora del 8.96% respecto al mejor MAE corto plazo del experimento de Variación de Cantidad de Datos.
- Horizonte largo plazo: El mejor modelo en este experimento fue ANN con MAE Validación = 0.004791. Esto representa una mejora del 7.42% respecto al mejor MAE largo plazo del experimento de Variación de Cantidad de Datos.

La adición de indicadores fundamentales también mejoró el MAE del mejor modelo en ambos horizontes, aunque la mejora porcentual fue la menor entre los cuatro grupos de características probados, especialmente para corto plazo.

12.2.7.5. COMPARATIVA RESULTADOS MAE BASE VS ADICIÓN TODAS LAS CARACTERÍSTICAS

Adición	Modelo	Frecuencia datos	MAE Validación	MAE Base	Variación MAE
Todos	ann	1h	0,000120	0,000172	-30,23%
Todos	ann	1d	0,004421	0,005175	-14,57%
Todos	cnn	1h	0,000126	0,000134	-5,97%
Todos	cnn	1d	0,005405	0,005282	2,33%
Todos	lstm	1h	0,001380	0,000669	106,28%
Todos	lstm	1d	0,005305	0,006235	-14,92%
Todos	transformer	1h	0,001218	0,000738	65,04%
Todos	transformer	1d	0,006240	0,006788	-8,07%

Tabla 27 - Comparativa de resultados MAE base vs MAE todas las características

Se puede evidenciar en la Tabla 27 que los resultados para CNN que fue el mejor modelo de predicción sin características adicionales en el corto plazo, mejoran reduciendo el error con la adición de esta característica en un 5.97 %. En este caso se evidencia una mejora también del ANN, reduciendo el MAE un 30.23 % y siendo el modelo con el mejor desempeño para esta métrica. Sin embargo, el modelo de Transformers, con la adición de todas las características se ve afectado, dado que aumenta su error en un 106.28 %, lo cual disminuye su desempeño. El mejor modelo para la predicción de un día fue el de ANN, y con la adición de indicadores fundamentales este modelo tiende a mejorar su efectividad sobre la predicción, reduciendo su error en un 14.57%.

12.2.7.6. DESEMPEÑO POR HORIZONTE Y MODELO:

Una observación consistente a lo largo de todos los sub-experimentos es la clara diferencia en la idoneidad de los modelos según el horizonte de predicción.

- Para el **corto plazo (horario)**, el modelo **CNN** demostró ser superior en la mayoría de los escenarios al añadir características individuales (técnicas, estacionales, alta frecuencia, fundamentales), logrando los menores errores (MAE) y los ajustes más altos (R^2). Sin embargo, cuando se combinaron **todas las características**, el modelo **ANN** logró superarlo ligeramente en términos de MAE, sugiriendo que ANN puede integrar información diversa de manera más efectiva a corto plazo si se le proporciona un conjunto rico de características. Los modelos LSTM y Transformer consistentemente mostraron un desempeño inferior en corto plazo.
- Para el **largo plazo (24 horas)**, el modelo **ANN** fue el claro ganador en todos los escenarios, logrando el menor MAE, el R^2 más alto y el SNR más elevado (menor incertidumbre) en el conjunto de validación. Su robustez, evidenciada por caídas de rendimiento relativamente menores entre entrenamiento y validación/test en comparación con otros modelos (especialmente en R^2), lo posiciona como la arquitectura más adecuada para predicciones a 6 días con los datos y configuraciones probadas. CNN mostró ser competitivo, pero a menudo con indicios de mayor sobreajuste o inestabilidad (como en el SNR con datos fundamentales). LSTM y Transformer tuvieron dificultades significativas en largo plazo, con errores más altos y ajustes R^2 inferiores.

12.2.7.7. IMPACTO DE LAS CARACTERÍSTICAS ADICIONALES

La adición de características adicionales, en general, demostró ser beneficiosa, mejorando la precisión del mejor modelo de cada horizonte en comparación con la línea base del Experimento de Variación de Cantidad de Datos. (que solo usaba OHLC y derivados).

- Para **corto plazo**, los **datos estacionales** proporcionaron la mayor mejora porcentual en el MAE del modelo CNN (-38.81%), seguidos de cerca por los **indicadores técnicos** (-30.60%). Los datos de alta frecuencia también aportaron valor (-24.63%), mientras que los indicadores fundamentales tuvieron el menor impacto positivo (-8.96%). Esto sugiere que los patrones cíclicos y las señales técnicas de corto plazo son cruciales para las predicciones a 6 horas.
- Para **largo plazo**, la **combinación de todas las características** resultó ser la más efectiva para el modelo ANN, logrando la mayor reducción porcentual del MAE (-14.61%) respecto al Experimento de Variación de Cantidad de Datos. Individualmente, los **indicadores técnicos** (-9.58%) fueron los más beneficiosos para ANN, seguidos por los datos de alta frecuencia (-8.31%) y los fundamentales (-7.42%). Los datos estacionales aportaron la menor mejora (-6.47%) para ANN en este horizonte. Esto indica que, para predicciones a largo plazo, una visión más holística que combine diferentes tipos de información es ventajosa, especialmente para el modelo ANN.

El análisis del SNR confirmó que la incertidumbre de las predicciones aumenta considerablemente con el horizonte temporal. Para todos los modelos y conjuntos de características, el SNR en largo plazo fue drásticamente menor (típicamente una reducción del 75-80%) que en corto plazo. Los modelos ANN y CNN consistentemente mostraron los SNR más altos en validación para corto plazo, indicando predicciones más seguras. Para largo plazo, ANN mantuvo el SNR más alto, alineado con su superioridad en precisión. Es interesante notar que los datos de alta frecuencia y la combinación de todas las características parecieron mitigar ligeramente la caída del SNR para el modelo Transformer en largo plazo, aunque no se tradujo en la mejor precisión absoluta.

12.3. EXPERIMENTO DE USO DE EXTRACTOR DE CARACTERÍSTICAS CON AUTOENCODER

En el presente experimento del proyecto, se procede a investigar el efecto de anteponer distintas arquitecturas de Aprendizaje Profundo, las cuales fungen como **extractores de características** pre-entrenados con un autoencoder, a un **predictor común y fijo**. El objetivo fundamental es determinar si este pre-entrenamiento del extractor de características con autoencoder, representa una mejora en el desempeño predictivo final, de esta forma encontrar qué tipo de extractor (ANN, CNN, LSTM o Transformer, cada uno configurado con dos capas) resulta ser el más beneficioso.

El **predictor fijo** utilizado en todos los experimentos de este experimento consiste en: dos capas Convolucionales 1D (Conv1D), seguidas de una capa LSTM y finalizando con una capa Bayesiana con Dense bias para la predicción y estimación de incertidumbre.

Para la ejecución de estos experimentos, se emplea el conjunto de datos base (compuesto por OHLC y derivados simples, cuya descripción se encuentra en la Sección 11.1), al cual se suman las 11 características provenientes de la descomposición de series de tiempo (STL, Wavelet y Multitaper, detalladas en la Sección 12.1). Se mantiene constante la cantidad de datos de entrenamiento en 25,200 registros.

Este experimento se articula en dos sub-experimentos principales:

- **Sección 14.1 (Extractores sin Pre-entrenamiento):** Se evalúan cuatro tipos de extractores (ANN, CNN, LSTM, Transformer). La salida de cada extractor se pasa luego al predictor fijo.
- **Sección 14.2 (Extractores con Pre-entrenamiento mediante Autoencoder):** Se replica la metodología implementada en la Sección 14.1, con la distinción fundamental de que los módulos extractores son pre-entrenados utilizando una arquitectura de autoencoder. Posteriormente, estos extractores, ya con sus pesos inicializados mediante pre-entrenamiento, se integran al predictor fijo con el fin de evaluar si esta estrategia mejora la capacidad de extracción de características y también, el rendimiento de la predicción.

Nota sobre la Métrica SNR: Para la evaluación de la incertidumbre de las predicciones, se mantiene la definición de la Relación Señal/Ruido (SNR) empleada en secciones previas. Esta métrica se calcula como el cociente entre la media de las predicciones generadas por la capa Bayesiana final y su desviación estándar, lo que representa una medida inversa de la incertidumbre relativa inherente a la predicción.

12.3.1. EXTRACTORES SIN PRE-ENTRENAMIENTO

En esta sección, se presentan los resultados de combinar cada uno de los cuatro tipos de extractores (ANN, CNN, LSTM y Transformer, cada uno con dos capas) con el predictor fijo. Los extractores reciben las 12 características de entrada base.

Tipo de Extractor	Horizonte	MAE Train	MAE Test	MAE Validación	Variación Train - Validación	Variación Test-Validación
ANN	Corto Plazo	0.000148	0.000093	0.000106	-28.38%	+13.98%
ANN	Largo Plazo	0.005002	0.003077	0.004411	-11.82%	+43.35%
CNN	Corto Plazo	0.000089	0.000049	0.000062	-30.34%	+26.53%
CNN	Largo Plazo	0.004964	0.002926	0.004312	-13.13%	+47.37%
LSTM	Corto Plazo	0.000252	0.000147	0.000198	-21.43%	+34.69%
LSTM	Largo Plazo	0.004730	0.003382	0.004479	-5.31%	+32.44%
Transformer	Corto Plazo	0.000181	0.000148	0.000140	-22.65%	-5.41%
Transformer	Largo Plazo	0.005089	0.003001	0.004416	-13.22%	+47.15%

Tabla 28 - Métricas de Desempeño (MAE) - Extractores sin Pre-entrenamiento

Análisis de Resultados (MAE) - Extractores sin Pre-entrenamiento:

En la Tabla 28 se presentan los resultados del Error Absoluto Medio (MAE). Para el horizonte de predicción a corto plazo, se observa que el Extractor CNN (MAE Val = 0.000062) permite al predictor fijo alcanzar la mayor precisión, seguido por el Extractor ANN (MAE Val = 0.000106). Los extractores LSTM y Transformer resultan en errores de validación más elevados.

Para el horizonte de largo plazo, el Extractor CNN (MAE Val = 0.004312) nuevamente se destaca como el más efectivo, proporcionando el menor MAE de validación. Le siguen de cerca los extractores ANN (MAE Val = 0.004411) y Transformer (MAE Val = 0.004416). Como es de esperar, el error de predicción es considerablemente mayor en el horizonte a largo plazo en comparación con el horizonte a corto plazo para todas las combinaciones de extractor.

Tipo de Extractor	Horizonte	R ² Train	R ² Test	R ² Validación	Variación Train-Validación	Variación Test-Validación
ANN	Corto Plazo	0.999996	0.999920	0.999974	0.00%	+0.01%
ANN	Largo Plazo	0.996395	0.915423	0.960263	-3.63%	+4.90%
CNN	Corto Plazo	0.999998	0.999976	0.999990	0.00%	0.00%
CNN	Largo Plazo	0.996434	0.923713	0.961650	-3.49%	+4.11%
LSTM	Corto Plazo	0.999987	0.999744	0.999895	-0.01%	+0.02%
LSTM	Largo Plazo	0.996831	0.900237	0.958664	-3.83%	+6.49%
Transformer	Corto Plazo	0.999993	0.999824	0.999953	0.00%	+0.01%
Transformer	Largo Plazo	0.996251	0.920007	0.959683	-3.67%	+4.31%

Tabla 29 - Métricas de Desempeño (R²) - Extractores sin Pre-entrenamiento

Análisis de Resultados (R²) - Extractores sin Pre-entrenamiento:

La Tabla 29 detalla el Coeficiente de Determinación (R²). En el horizonte a corto plazo se evidencia que todos los extractores permiten al predictor fijo alcanzar valores de R² excepcionalmente altos en validación (superiores a 0.9998), indicando un ajuste casi perfecto. El Extractor CNN (R² Val = 0.999990) resulta en el valor más elevado.

Para el horizonte a largo plazo, los valores de R² en validación disminuyen. El Extractor CNN (R² Val = 0.961650) nuevamente proporciona el mejor R², seguido de cerca por el Extractor ANN. Se observa una caída en el R² entre los conjuntos de entrenamiento y test para todos los extractores en este horizonte, aunque con una recuperación parcial en el conjunto de validación.

Tipo de Extractor	Horizonte	SNR Train	SNR Test	SNR Validación	Variación Train-Validación	Variación Test-Validación
ANN	Corto Plazo	6,299.94	8,518.65	6,753.56	+7.20%	-20.72%
ANN	Largo Plazo	1,468.63	1,948.43	1,639.11	+11.61%	-15.88%
CNN	Corto Plazo	6,455.24	9,291.22	7,075.01	+9.60%	-23.85%
CNN	Largo Plazo	1,162.26	1,552.35	1,302.67	+12.08%	-16.08%
LSTM	Corto Plazo	6,086.07	8,396.41	6,736.36	+10.68%	-19.77%
LSTM	Largo Plazo	1,282.23	1,797.29	1,536.88	+19.86%	-14.49%
Transformer	Corto Plazo	5,136.50	6,272.24	5,498.02	+7.04%	-12.34%
Transformer	Largo Plazo	1,210.80	1,611.35	1,339.17	+10.60%	-16.89%

Tabla 30 - Métricas de Desempeño (SNR) - Extractores sin Pre-entrenamiento

Análisis de Resultados (SNR) - Extractores sin Pre-entrenamiento:

En la Tabla 30 se presentan los valores de SNR. Para el horizonte a corto plazo el Extractor CNN (SNR Val = 7075.01) permite al predictor alcanzar la menor incertidumbre relativa. Los extractores ANN y LSTM también conducen a valores altos de SNR.

Al extender el horizonte a largo plazo, se observa una disminución considerable del SNR para todas las combinaciones, indicativo de un incremento en la incertidumbre. En este horizonte, el Extractor ANN (SNR Val = 1639.11) es el que resulta en la menor incertidumbre relativa para el predictor fijo.

12.3.2. MEJOR PREDICCIÓN PARA EXTRACTORES SIN PRE-ENTRENAMIENTO

Basándose en los resultados, el extractor **CNN** proporcionó la mejor predicción para el horizonte a corto plazo (MAE=0.000062) como para el horizonte a largo plazo (MAE=0.004312).

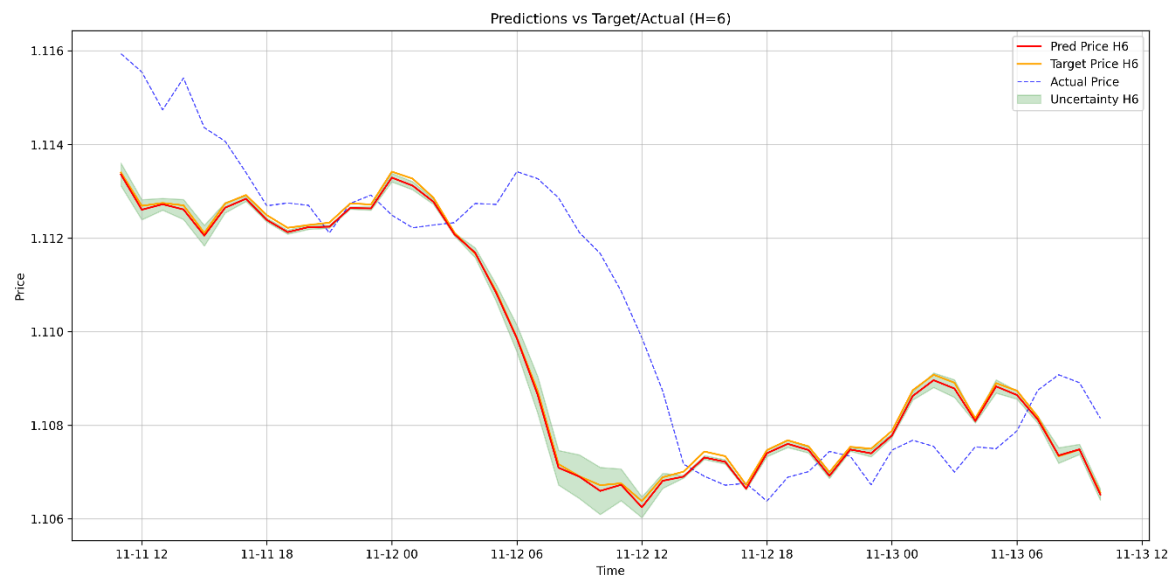


Figura 25 - Comparación real vs predicho (corto plazo) usando la mejor combinación (Extractor ANN + Modelo Predictivo CNN).

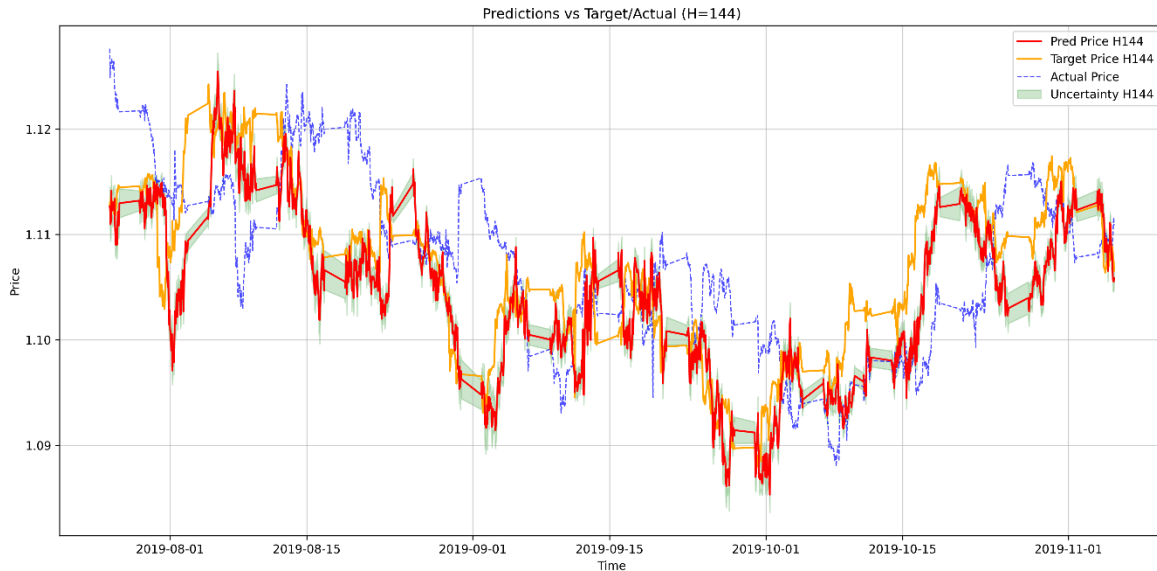


Figura 26 - Comparación real vs predicho (largo plazo) usando la mejor combinación (actualmente Extractor ANN + Modelo Predictivo CNN).

Como se observa en las Figuras 25 y 26, tanto las predicciones de corto plazo como las de largo plazo se ajustan a la señal de Target con un error menor al de todas las predicciones de secciones anteriores.

12.3.3. EXTRACTORES CON PRE-ENTRENAMIENTO (AUTOENCODER)

En esta sección, se investiga si pre-entrenar los módulos extractores (ANN, CNN, LSTM, Transformer de dos capas) utilizando un autoencoder mejora el rendimiento final al combinarlos con el predictor fijo.

12.3.3.1. METODOLOGÍA Y ENTRENAMIENTO DEL AUTOENCODER

Se utilizó un autoencoder para aprender representaciones de los datos de entrada. La arquitectura del autoencoder consistió en un encoder y un decoder, cuyas arquitecturas específicas variaron según el tipo de extractor que se buscaba pre-entrenar (ANN, CNN, LSTM, Transformer).

El autoencoder se entrenó con los mismos hiperparámetros que los predictores, incluyendo una función de pérdida $loss = Huber + Sigma * MMD$.

Para el entrenamiento en reconstruir la señal de entrada del encoder en la salida del decoder, se usaron 1000 épocas, donde sigma es 0.01.

- El autoencoder se entrenó con un dataset de 25200 datos de entrenamiento con fechas previas a las de los datos usados durante el entrenamiento de los predictores.
- **Datos para el Autoencoder:** El conjunto de datos total se dividió en 6 partes (d1 a d6). Los tres primeros (d1, d2, d3) se utilizaron para el entrenamiento, validación y prueba del autoencoder,

respectivamente. El autoencoder se entrenó con 25,200 registros (d1) y se validó con 6,300 registros (d2). No se utilizó un conjunto de prueba explícito (d3) para la evaluación final del autoencoder en esta etapa, aunque estaba disponible.

- **Objetivo del Autoencoder:** Reconstruir la señal de entrada (compuesta por múltiples características) en la salida de su decoder. El error (MAE) se midió sobre esta tarea de reconstrucción de las aproximadamente 41 características de entrada.
- **Inicialización:** Una vez entrenado el autoencoder, la porción del encoder (guardada como `encoder.h5.keras`) se utilizó para inicializar las dos primeras capas del módulo extractor correspondiente. Posteriormente, el sistema completo (Extractor Pre-entrenado + Modelo Predictivo base) se entrenó de forma supervisada utilizando los conjuntos de datos d4, d5 y d6 para entrenamiento, validación y prueba del predictor, respectivamente.

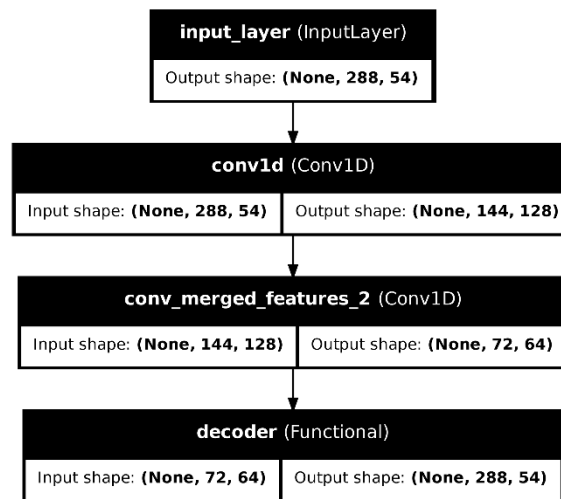


Figura 27 - Modelo del Autoencoder usando capas Conv1D, generado usando Keras.

En la Figura 27 se muestra el modelo del autoencoder, se puede ver las capas del encoder (dos Conv1D), y el decoder que tiene las capas del mismo tamaño que las capas Conv1D pero de tipo Conv1DTranspose, usando el mismo `stride=2` y `kernel_size=3` que aumenta la dimensionalidad al doble en lugar de disminuirla a la mitad como hace en las capas Conv1D.

Se puede observar la **reducción de dimensionalidad** de (288,54) a (72,64) en el encoder, mientras el decoder muestra un incremento de dimensionalidad de (72,64) a (288,54).

El Decoder fue definido como una función Lambda (función anónima), con dos capas Conv1DTranspose cuyo propósito es recomponer la señal de entrada del Encoder en su salida.

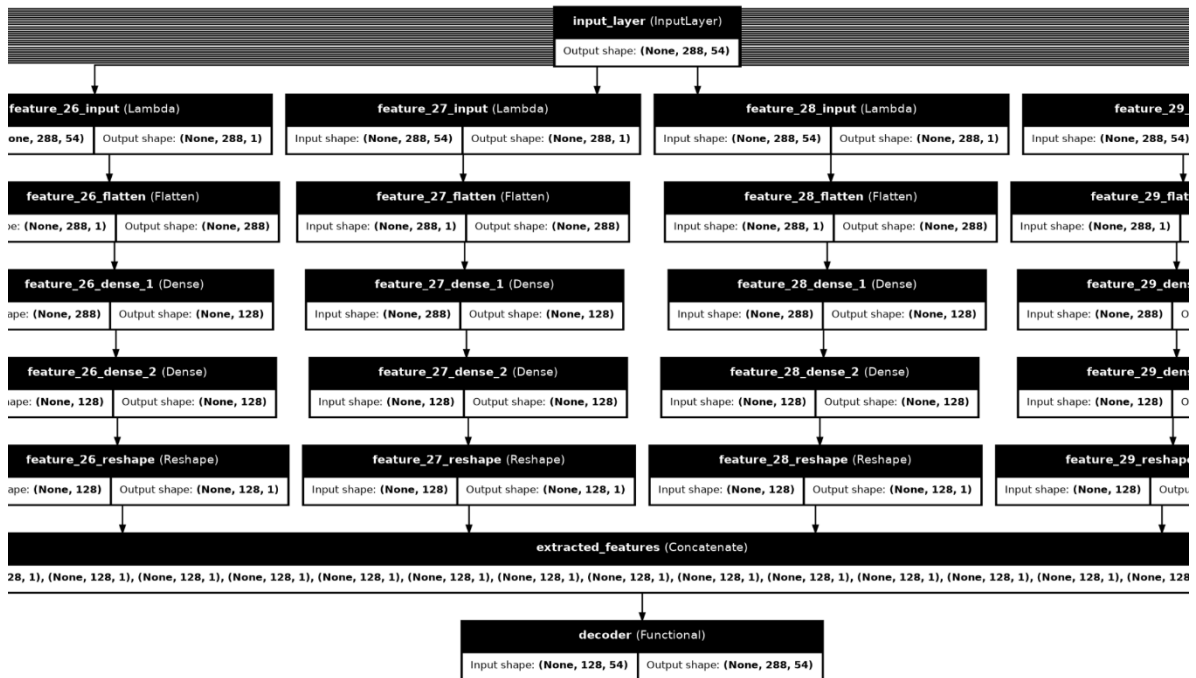


Figura 28 – Fragmento del Modelo del Autoencoder usando capas Dense (ANN) en paralelo para cada característica de entrada, plot generado usando Keras.

La misma estructura se usó para los otros modelos de autoencoder(ANN, LSTM, Transformers), usando dos capas de cada modelo y para el decoder sus capas inversas, en el caso de ANN se usan dos capas densas por característica en paralelo como se muestra en la Figura 28 (el diagrama del modelo completo se encuentra en el **Anexo F**).

Mientras que los modelos de autoencoder de LSTM, Transformers y CNN están entre 160.000 y 200.000 parámetros, el modelo de ANN con procesamiento paralelo de características tiene más de 1.000.000 de parámetros y su entrenamiento es muy lento.

12.3.3.2. RESULTADOS DEL ENTRENAMIENTO DE LOS AUTOENCODERS

Antes de evaluar el impacto en la predicción final, se analizó el rendimiento de los propios autoencoders durante su fase de entrenamiento de reconstrucción. Los resultados (MAE de reconstrucción) para cada tipo de arquitectura base del autoencoder fueron los siguientes:

Arquitectura Base del Autoencoder	Horizonte	MAE Train (Recon.)	MAE Validación (Recon.)	Diferencia (%) Train vs Val
ANN	Corto Plazo	0.0737	0.4504	+510.9%
ANN	Largo Plazo	0.0892	0.1052	+17.9%
CNN	Corto Plazo	0.0168	0.0178	+5.9%
CNN	Largo Plazo	0.0073	0.0067	-8.4%
LSTM	Corto Plazo	0.0294	0.0335	+13.8%
LSTM	Largo Plazo	0.0148	0.0161	+9.1%
Transformer	Corto Plazo	0.0268	0.0276	+3.0%
Transformer	Largo Plazo	0.0228	0.0242	+6.3%

Tabla 31 - MAE de Reconstrucción - Entrenamiento de Autoencoders

Análisis del Entrenamiento del Autoencoder: La Tabla 31 muestra el error de reconstrucción (MAE) de los autoencoders para ambos horizontes (corto plazo y largo plazo). Es crucial notar que este MAE **no es comparable** al MAE de predicción de las secciones anteriores, ya que mide la capacidad del autoencoder para replicar su entrada, no para predecir valores futuros. Para **corto plazo**, el autoencoder basado en **ANN** muestra un MAE de validación extremadamente alto (0.4504) y una diferencia enorme respecto al entrenamiento (+510.9%), sugiriendo falta de ajuste **o sobreajuste masivo**. Los extractores basados en **CNN** y **Transformer** tienen los MAE más bajos (0.0178 y 0.0276) y diferencias mínimas Train-Val (+5.9% y +3.0%), indicando **buena convergencia del error a un valor similar en train que en validation y por tanto buena generalización**. LSTM (0.0335) es intermedio. Para **largo plazo**, el autoencoder **CNN** logra el MAE de validación más bajo (0.0067) y una diferencia Train-Val negativa (-8.4%), indicando excelente convergencia y generalización. LSTM (0.0161) y Transformer (0.0242) también convergen bien. Sin embargo, el autoencoder **ANN** (0.1052), aunque mejora respecto a corto plazo, sigue mostrando un error considerablemente mayor y una diferencia Train-Val (+17.9%) más alta que los demás (excepto LSTM corto plazo), lo que sigue poniendo en duda la calidad de las representaciones aprendidas por el encoder ANN, especialmente para corto plazo, lo cual puede tener como causa el **alto número de parámetros a entrenar en el autoencoder basado en ANN** ya que el procesamiento paralelo de las características eleva el número de parámetros de aproximadamente 160.000 en promedio en los modelos usados hasta ahora respecto a más de **1.000.000 de parámetros** del autoencoder de ANN, que requirió una GPU con 16GB de RAM para entrenar debido a que una GPU de 8GB mostraba errores de Out Of Memory(OOM). Por tanto, puede que se requiera más entrenamiento para un modelo tan grande ya que con las mismas épocas de entrenamiento que los demás modelos, como en este caso, puede estar en desventaja.

12.3.3.3. RESULTADOS DE PREDICCIÓN CON EXTRACTORES PRE-ENTRENADOS CON AUTOENCODER

En esta sección se presentan los resultados de utilizar los extractores pre-entrenados para cada modelo junto con cabezales predictivos fijos para todas las pruebas.

Resultados con Extractores Pre-entrenados: Las siguientes tablas muestran los resultados de los 4 modelos utilizados como extractores de características pre-entrenados y utilizados en conjunto con un predictor fijo.

Tipo de Extractor	Horizonte	MAE Train	MAE Test	MAE Validación	Variación Train-Validación	Variación Test-Validación
ANN	Corto Plazo	0.000299	0.000135	0.000110	-171.82%	-22.73%
ANN	Largo Plazo	0.005106	0.003026	0.004373	-16.76%	+30.80%
CNN	Corto Plazo	0.000072	0.000071	0.000076	+5.26%	+6.58%
CNN	Largo Plazo	0.004636	0.003970	0.004786	+3.13%	+17.05%
LSTM	Corto Plazo	0.000093	0.000066	0.000084	-10.71%	+21.43%
LSTM	Largo Plazo	0.004395	0.003298	0.004708	+6.65%	+29.95%
Transformer	Corto Plazo	0.000068	0.000042	0.000061	-11.48%	+31.15%
Transformer	Largo Plazo	0.004815	0.003461	0.004562	-5.55%	+24.13%

Tabla 32 - Métricas (MAE) - Extractores Pre-entrenados

Análisis de Resultados (MAE): La Tabla 32 muestra el MAE usando el extractor Transformer pre-entrenado. Para **corto plazo**, el modelo predictivo **Transformer** (MAE Val = 0.000061) obtiene el error más bajo, seguido por CNN (0.000076). Para **largo plazo**, el modelo predictivo **ANN** (MAE Val = 0.004373) es el más preciso.

Tipo de Extractor	Horizonte	R ² Train	R ² Test	R ² Validación	Variación Train-Validación	Variación Test-Validación
ANN	Corto Plazo	0.999983	0.999813	0.999966	-0.00%	-0.02%
ANN	Largo Plazo	0.996288	0.917764	0.960987	-3.67%	+4.50%
CNN	Corto Plazo	0.999999	0.999957	0.999986	0.00%	-0.00%
CNN	Largo Plazo	0.996979	0.869153	0.953594	-4.55%	+8.86%
LSTM	Corto Plazo	0.999998	0.999954	0.999981	0.00%	-0.00%
LSTM	Largo Plazo	0.997337	0.906805	0.953626	-4.58%	+4.91%
Transformer	Corto Plazo	0.999999	0.999979	0.999988	0.00%	-0.00%
Transformer	Largo Plazo	0.996904	0.896771	0.956144	-4.26%	+6.21%

Tabla 33 - Métricas (R²) - Extractores Pre-entrenados

Análisis de Resultados (R²): La Tabla 33 muestra el R² al usar el Extractor Transformer pre-entrenado. Para corto plazo, todos los modelos predictivos alcanzan R² de validación muy altos (>0.9999), siendo el CNN (0.999986) el que mejor se ajusta. Para largo plazo, el modelo predictivo ANN (R² Val = 0.960987) presenta el mejor ajuste, aunque todos los modelos muestran una caída considerable entre Train y Test, recuperándose parcialmente en Validación.

Tipo de Extractor	Horizonte	SNR Train	SNR Test	SNR Validación	Variación Train-Validación	Variación Test-Validación
ANN	Corto Plazo	6,983.02	8,912.22	7,491.78	-6.79%	+18.96%
ANN	Largo Plazo	1,484.70	1,882.25	1,658.74	-10.49%	+13.47%
CNN	Corto Plazo	6,300.41	9,028.52	6,796.60	-7.30%	+32.84%
CNN	Largo Plazo	849.01	1,046.01	928.22	-8.53%	+12.69%
LSTM	Corto Plazo	6,260.07	9,017.74	6,938.35	-9.78%	+29.97%
LSTM	Largo Plazo	796.64	958.38	911.08	-12.56%	+5.19%
Transformer	Corto Plazo	6,202.19	8,836.86	6,839.12	-9.31%	+29.21%
Transformer	Largo Plazo	1,045.52	1,340.80	1,201.11	-12.95%	+11.63%

Tabla 34 - Métricas (SNR) - Extractores Pre-entrenados

Análisis de Resultados (SNR): Para **corto plazo**, el modelo predictivo **ANN** (SNR Val = 7491.78) presenta la menor incertidumbre relativa, seguido de cerca por LSTM y Transformer. Para **largo plazo**, el modelo predictivo **ANN** (SNR Val = 1658.74) vuelve a ser el mejor. La caída del SNR al pasar de corto plazo a largo plazo es significativa para todos los modelos (ej., ANN: ~78%), indicando mayor incertidumbre en predicciones a largo plazo.

Mejor Predicción General (Extractores Pre-entrenados con autoencoder): Para corto plazo, la combinación de extractor de características usando modelo de Transformers Pre-entrenado con Autoencoder (MAE=0.000061) fue la más precisa.

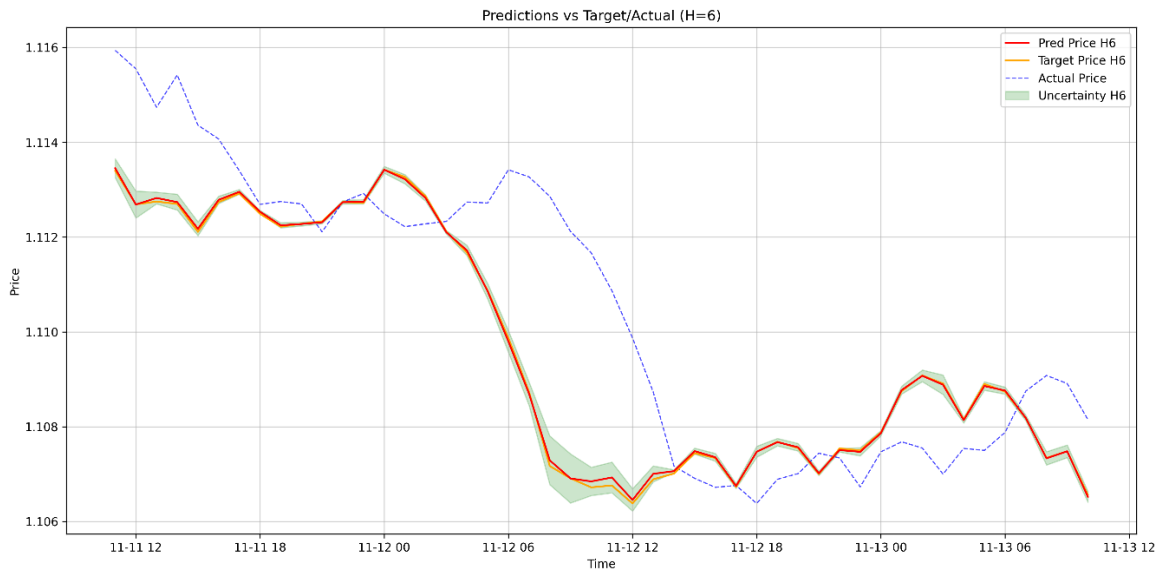


Figura 29 - Comparación real vs predicho (corto plazo) usando la mejor combinación con extractor pre-entrenado.

En la Figura 29 se observa un excelente ajuste de la predicción al valor deseado y una baja incertidumbre.

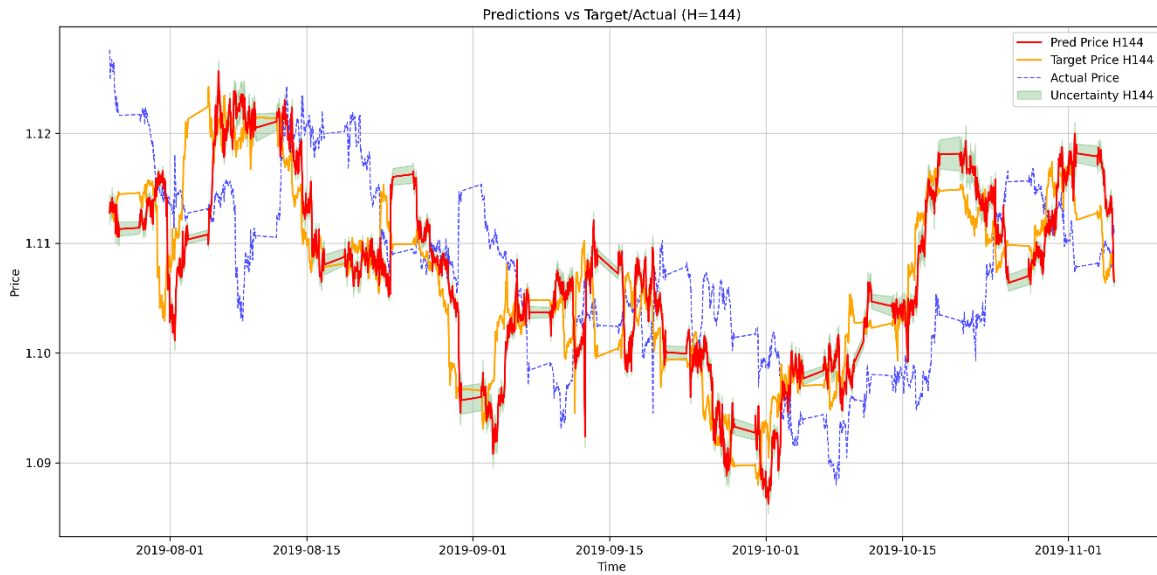


Figura 30 - Comparación real vs predicho (largo plazo) usando la mejor combinación con extractor pre-entrenado.

En la Figura 30 se muestra la mejor predicción a 6 días hasta este punto, a pesar de que presenta mayor ruido que la predicción de corto plazo, puede ser útil en la estrategia de trading ya que predice correctamente la mayor parte de los movimientos del precio real.

12.4. EXPERIMENTOS DE USO DE DATOS SINTÉTICOS PARA ENTRENAMIENTO

Para este experimento del proyecto se implementó un proceso para ampliar el conjunto de datos original mediante la generación de datos sintéticos, con el objetivo de evaluar el impacto de diferentes volúmenes y tipos de datos en el desempeño de los modelos de predicción de la tasa de cambio EUR/USD. Para ello, se diseñaron y desarrollaron tres generadores de datos sintéticos que operan sobre la columna "close" de los datos de entrenamiento:

12.4.1. EXPERIMENTO DE GENERACIÓN DE DATOS SINTÉTICOS CON VAE:

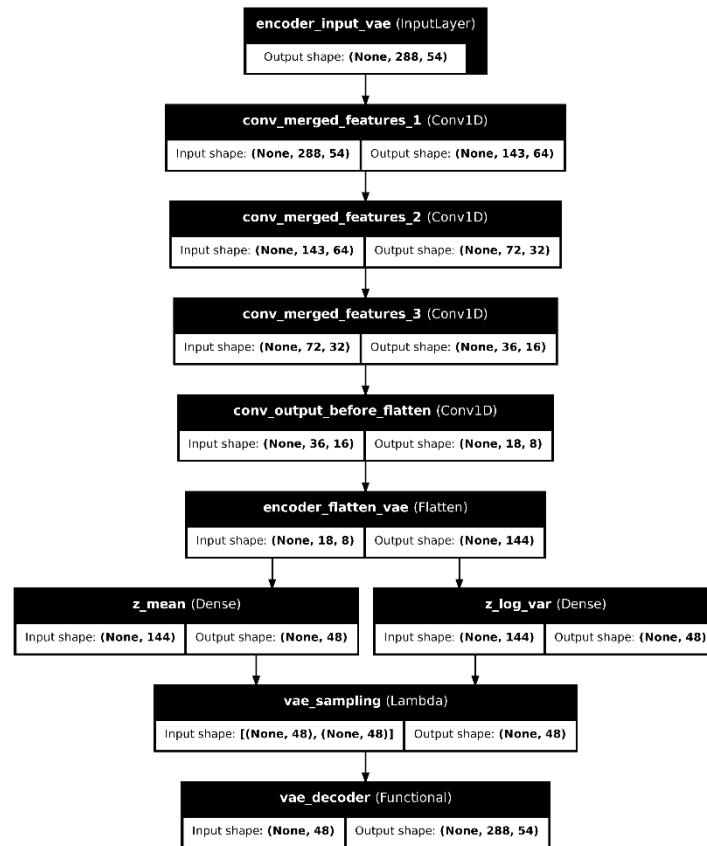


Figura 31 – Plot de modelo de VAE

Generador VAE: Se configuró un Autoencoder Variacional (VAE) similar al usado en la sección 14 pero compuesto por un codificador con 3 capas Conv1D con stride 2, con la adición de una capa Estadística (o estocástica, no confundir con la capa Bayesiana) que estima con dos capas dense paralelas, la media (z_mean) y la desviación estándar (z_log_var) de cada característica del espacio latente como se muestra en la Figura 31 y estas entradas se pasan al decodificador en lugar del espacio latente (salida del encoder), además el decoder tiene una capa densa de entrada y una capa de reshape que permiten reconstruir la estructura 3D del espacio latente seguido por 3 capas conv1DTranspose. Esta configuración permite codificar la información de la serie temporal en un espacio latente, **se calcula la distribución de cada característica del espacio latente** y posteriormente decodifica para reconstruir la señal de entrada.

Este modelo se implementó en el programa desarrollado para el proyecto en el repositorio synthetic-datagen, y una vez entrenado, se utilizó para generar datasets sintéticos que equivalen a 2 años, 4 años y 8 años, usando en todos los casos 4 años de datos reales para entrenamiento y solo un año de datos reales para validación y test. Así, estos datos generados se integraron concatenando las filas del dataset generado antes del dataset de datos reales del conjunto de entrenamiento original, permitiendo realizar pruebas y evaluaciones comparativas de los modelos

de deep learning (incluyendo arquitecturas como CNN, LSTM y Transformers) en condiciones de mayor volumen de datos. De este modo, se pudo analizar el efecto del aumento de datos sintéticos en la robustez y precisión de las predicciones de la tasa de cambio EUR/USD.

El cálculo del promedio y desviación estándar por característica del espacio latente durante training e inferencia se concatena con las características del espacio latente.

Para generar los datos sintéticos se muestrean datos aleatorios con distribución caracterizada por la media y desviación estándar por característica del espacio latente que se tienen calculados y se alimenta en estos valores en la entrada del decoder para producir en su salida los datos sintéticos con el mismo número de las características que el dataset original.

Este método de generación no es el más apto para series de tiempo ya que se deben asignar las fechas a los datos sintéticos en lugar de que los datos sintéticos se generen dadas las fechas para que estos se generen como ocurre con un Conditional VAE (sección 15.2) especialmente si se realimenta con los datos sintéticos generados anteriormente para generación secuencial.

Debido a nuestras limitaciones en capacidad computacional principalmente y también limitaciones de tiempo, ya que el entrenamiento del VAE fue extremadamente lento por la complejidad de los modelos y la cantidad de datos, por tanto, no se realizaron pruebas GAN o similares modelos aptos para series de tiempo, pero la implementación de un VAE simple, sirve de introducción para explorar el campo de la generación de datos financieros con el hardware disponible (RTX 4090m con 16GB y RTX4070m con 8GB de GPU RAM), para contrastar con el Conditional VAE y como base para trabajo futuro en series de tiempo, como una CVAE-GAN.

Como datos de validación y test, se usa un año de datos reales posteriores a los datos combinados de training (reales + sintéticos) para cada uno.

12.4.1.1. ANÁLISIS DE MAE PARA CORTO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
transformer	12600	Serie de tiempo	0,000411	0,000172	0,000221	-46,23%	28,49%
transformer	25200	Serie de tiempo	0,000535	0,000267	0,000255	-52,34%	-4,49%
transformer	50400	Serie de tiempo	0,000452	0,000186	0,000225	-50,22%	20,97%

Tabla 35 - Resultados MAE de experimento de uso de datos sintéticos para corto plazo

En la Tabla 35 se muestran los resultados del MAE del entrenamiento del modelo de transformers usando datos sintéticos, ya que este modelo fue el mejor en la sección 14. Se revelan valores consistentemente bajos en los conjuntos de entrenamiento, prueba y validación, lo que indica una buena precisión en las predicciones del modelo. Los errores en el conjunto de entrenamiento se mantienen bajos a través de las diferentes cantidades de datos, sugiriendo una capacidad de aprendizaje estable. El rendimiento en el conjunto de validación también es favorable, indicando una adecuada generalización a datos no vistos durante el entrenamiento. Sin embargo, se observa una variabilidad significativa en la diferencia porcentual del MAE entre el conjunto de prueba y el de validación, con fluctuaciones considerables que sugieren una sensibilidad del modelo al conjunto de datos de prueba específico.

Esta variabilidad y los errores significativos respecto a los demás experimentos de este proyecto podrían deberse a que se generan datos episódicos, sin estructura temporal definida y sin condiciones que generen los datos basados en la fecha o datos pasados como se hace en el Conditional VAE de la sección 15.2.

12.4.1.2. ANÁLISIS DE R2 PARA CORTO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
transformer	12600	Serie de tiempo	0,999991	0,999737	0,999889	-0,01%	0,02%
transformer	25200	Serie de tiempo	0,999986	0,999419	0,999855	-0,01%	0,04%
transformer	50400	Serie de tiempo	0,999990	0,999704	0,999889	-0,01%	0,02%

Tabla 36 - Resultados R2 de experimento de uso de datos sintéticos para corto plazo

La Tabla 36 muestra el análisis del coeficiente de determinación (R^2) para el modelo Transformer revela un rendimiento excepcionalmente alto y consistente en los conjuntos de entrenamiento, prueba y validación, con valores cercanos a 1. Esto indica que el modelo explica una proporción muy alta de la varianza en la variable objetivo para las tres particiones de los datos. La variación entre el rendimiento del entrenamiento y la validación es mínima, sugiriendo un ajuste adecuado sin un sobreajuste significativo. De manera similar, se pueden ver valores peores que los de los demás experimentos posiblemente por la falta de estructura secuencial de los datos generados.

12.4.1.3. ANÁLISIS DE SNR PARA CORTO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
transformer	12600	Serie de tiempo	1.409,42	9.030,70	8.131,52	476,94%	-9,96%
transformer	25200	Serie de tiempo	1.230,70	8.739,39	8.452,14	586,77%	-3,29%
transformer	50400	Serie de tiempo	1.235,12	8.209,17	8.283,17	570,64%	0,90%

Tabla 37 - Resultados SNR de experimento de uso de datos sintéticos para corto plazo

La Tabla 37 muestra los resultados de SNR para el modelo Transformer y revela valores significativamente más altos en los conjuntos de prueba y validación en comparación con el conjunto de entrenamiento. El SNR en el conjunto de entrenamiento se mantiene relativamente bajo a través de las diferentes cantidades de datos, lo que sugiere que el modelo podría estar luchando por distinguir la señal del ruido en los datos de entrenamiento. Sin embargo, el SNR en los conjuntos de prueba y validación es considerablemente mayor, indicando una mejor capacidad para identificar la señal relevante en datos no vistos. La gran diferencia positiva entre el rendimiento del entrenamiento y la validación sugiere un posible subajuste en el conjunto de entrenamiento. La variación relativamente pequeña en la diferencia del SNR entre el conjunto de prueba y el de validación indica una consistencia razonable en el rendimiento del modelo entre estos dos conjuntos de datos no vistos.

12.4.1.4. ANÁLISIS DE MAE PARA LARGO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
cnn	12600	Serie de tiempo	0,022113	0,004234	0,006145	-72,21%	45,13%
cnn	25200	Serie de tiempo	0,031744	0,004729	0,006702	-78,89%	41,72%
cnn	50400	Serie de tiempo	0,044428	0,006257	0,007046	-84,14%	12,61%

Tabla 38 - Resultados MAE de experimento de uso de datos sintéticos para largo plazo

La Tabla 38 muestra para el modelo CNN revela valores bajos en los conjuntos de prueba y validación en comparación con el conjunto de entrenamiento, lo que sugiere una tendencia al sobreajuste. El MAE en el conjunto de entrenamiento aumenta a medida que se incrementa la cantidad de datos, lo que podría indicar una mayor dificultad para modelar conjuntos de datos más grandes o la necesidad de ajustar la capacidad del modelo. A pesar de esto, el rendimiento en los conjuntos de prueba y validación se mantiene relativamente bajo, aunque con una variación considerable entre ellos, como se evidencia en la columna "Variación Test - Validación". Esta fluctuación sugiere que la capacidad de generalización del modelo a datos no vistos puede ser inconsistente y dependiente de las características específicas del conjunto de prueba. La gran diferencia negativa entre el rendimiento del entrenamiento y la validación refuerza la observación de un posible sobreajuste.

12.4.1.5. ANÁLISIS DE R2 PARA LARGO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
cnn	12600	Serie de tiempo	0,965003	0,845856	0,921650	-4,49%	8,96%
cnn	25200	Serie de tiempo	0,951337	0,812047	0,908833	-4,47%	11,92%
cnn	50400	Serie de tiempo	0,906425	0,687210	0,901083	-0,59%	31,12%

Tabla 39 - Resultados R2 de experimento de uso de datos sintéticos para largo plazo

La Tabla 39 muestra los resultados para el modelo CNN y revela un rendimiento generalmente alto en los conjuntos de entrenamiento, prueba y validación, indicando que el modelo explica una buena proporción de la varianza en la variable objetivo. Se observa una ligera disminución en el R^2 del conjunto de entrenamiento a medida que aumenta la cantidad de datos, lo que podría sugerir una mayor dificultad para modelar conjuntos de datos más grandes. Sin embargo, el R^2 en los conjuntos de prueba y validación se mantiene relativamente alto, aunque con una diferencia notable entre ellos, como se evidencia en la columna "Variación Test - Validación". Esta variación sugiere cierta inestabilidad en la capacidad de generalización del modelo a datos no vistos, posiblemente influenciada por las características específicas del conjunto de prueba. La pequeña diferencia negativa entre el rendimiento del entrenamiento y la validación indica un sobreajuste limitado.

12.4.1.6. ANÁLISIS DE SNR PARA LARGO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
cnn	12600	Serie de tiempo	287,14	619,15	481,56	67,71%	-22,22%
cnn	25200	Serie de tiempo	179,01	249,16	243,93	36,27%	-2,10%
cnn	50400	Serie de tiempo	273,26	318,83	334,65	22,47%	4,96%

Tabla 40 - Resultados SNR de experimento de uso de datos sintéticos para corto plazo

La Tabla 40 deja ver el SNR para el modelo CNN revela valores generalmente altos en los conjuntos de prueba y validación en comparación con el conjunto de entrenamiento. El SNR en el conjunto de entrenamiento disminuye significativamente a medida que se incrementa la cantidad de datos, lo que sugiere que el modelo podría estar capturando más ruido o patrones específicos del entrenamiento a medida que se expone a más información. Sin embargo, el SNR en los conjuntos de prueba y validación se mantiene considerablemente más alto, indicando una mejor capacidad para distinguir la señal del ruido en datos no vistos. La diferencia positiva entre el rendimiento del entrenamiento y la validación, que se hace más pronunciada con más datos, sugiere un posible subajuste en los conjuntos de entrenamiento más grandes. La variación en la diferencia del SNR entre el conjunto de prueba y el de validación indica cierta sensibilidad del modelo a las características específicas del conjunto de prueba.

12.4.2. EXPERIMENTO DE GENERACIÓN DE DATOS SINTÉTICOS CON SEQUENTIAL CONDITIONAL VAE

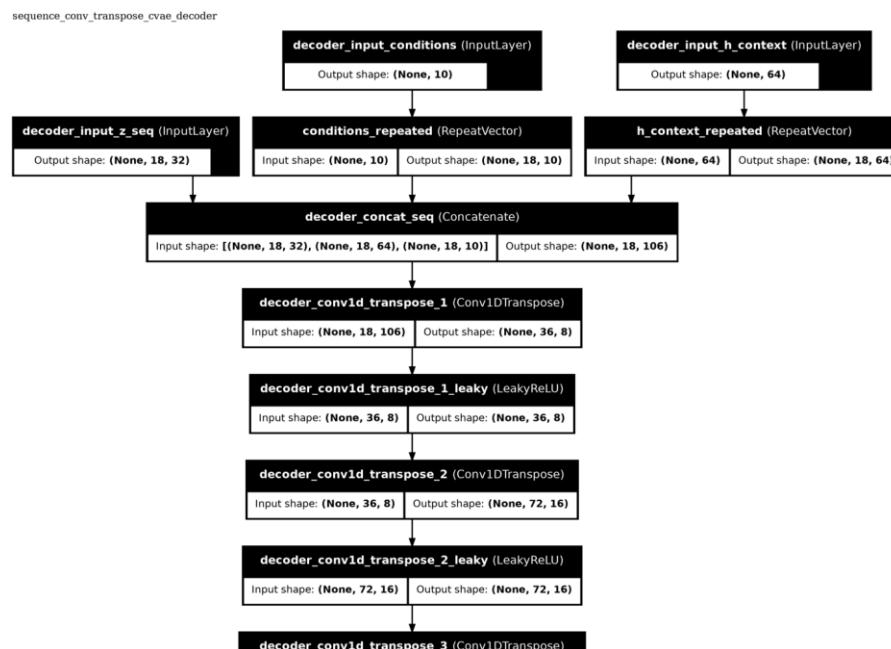


Figura 32 – Decodificador de Sequential Conditional VAE (Fragmento de plot generado con keras.utils.plot_model)

En esta fase del proyecto, se implementó un Autoencoder Variacional Condicional (CVAE), que representa una extensión del VAE previamente configurado. A diferencia del VAE estándar que

opera sin condicionamiento explícito, el CVAE fue entrenado para generar datos sintéticos de la serie temporal bajo la influencia de variables contextuales. Como se aprecia en la Figura 32 (el diagrama del modelo completo se encuentra en el **Anexo G**), el decodificador de este CVAE recibió como **entradas condicionales la fecha del tick que se estaba generando** (decoder_input_conditions), con el fin de modelar la estacionalidad inherente de los datos, y también **los últimos valores generados en el tick precedente**(decoder_input_h_context), adicionalmente a las salidas de la capa estadística concatenadas (decoder_input_z_seq) para capturar la **dependencia secuencial** de la serie temporal. También se puede apreciar que el decoder no tiene una capa de Reshape como en el VAE simple, sino que usa capas de repetición para adecuar las tres entradas para que tengan la misma dimensión temporal (tamaño: batch_size, dimensión temporal, num_features), con el propósito de concatenarlas directamente por el eje de características y usar esta concatenación tridimensional como entrada a las capas Conv1DTranspose que reconstruyen la señal de entrada del encoder en la salida del decoder.

Es crucial señalar que, en esta iteración, el CVAE se enfocó únicamente en la generación de características que no se derivan directamente de fórmulas predefinidas, como el precio de cierre, evitando la reconstrucción de indicadores técnicos que se calculan a partir de las características base generadas (24). Este enfoque difiere del VAE no condicional, que intenta generar todas las características de forma episódica, no secuencial, sin integrar una estructura temporal explícita mediante entradas condicionales, ni calcular características que dependen de otras como los indicadores técnicos. Las características que no se generaron se calcularon en base a la fecha (variables estacionales) y calculándolas desde los valores generados (indicadores técnicos) para obtener el mismo número de características que el dataset original.

Debido a limitaciones en capacidad computacional principalmente y también limitaciones de tiempo, ya que el entrenamiento del VAE fue extremadamente lento por la complejidad de los modelos y la cantidad de datos, no se realizaron pruebas de GAN o similares modelos aptos para series de tiempo, pero la implementación de un VAE condicional, sirve de introducción para explorar el campo de la generación de datos financieros con el hardware disponible (RTX 4090m con 16GB y RTX4070m con 8GB de GPU RAM), también sirve para contrastar con una VAE simple y como base para trabajo futuro en series de tiempo.

12.4.2.1. ANÁLISIS DE MAE PARA CORTO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
transformer	12600	Serie de tiempo	0,000099	0,000094	0,000128	29,29%	36,17%
transformer	25200	Serie de tiempo	0,000079	0,000075	0,000116	46,84%	54,67%
transformer	50400	Serie de tiempo	0,000058	0,000065	0,000101	74,14%	55,38%

Tabla 41 - Resultados MAE de experimento de uso de datos sintéticos para corto plazo

En la Tabla 41 se muestra análisis del Error Absoluto Medio (MAE) para el modelo Transformer en predicciones a corto plazo revela valores consistentemente bajos en los conjuntos de entrenamiento, prueba y validación, lo que indica una buena precisión en las predicciones del modelo. Los errores en el conjunto de entrenamiento disminuyen a medida que aumenta la cantidad de datos, sugiriendo una mejora en el aprendizaje con mayor exposición. El rendimiento

en los conjuntos de prueba y validación también es favorable, indicando una adecuada generalización a datos no vistos. Sin embargo, se observa una variabilidad considerable en la diferencia porcentual del MAE entre el conjunto de entrenamiento y validación, y aún más pronunciada entre el conjunto de prueba y validación. Estas variaciones significativas sugieren que la robustez del modelo podría ser sensible a las características específicas de las particiones de datos, a pesar de los valores absolutos bajos del error.

Se observa que el error en Test obtenido por el VAE condicional (0.00065) es peor que el obtenido en la sección 14 (MAE = 0.0004) y por tanto al adicionar datos sintéticos a los reales se empeora la predicción respecto a solo usar datos reales, a pesar de que mejora respecto al VAE simple.

Aunque se puede observar que a mayor número de datos sintéticos adicionados se reduce el MAE, lo que nos dice que existe la posibilidad que si se usa la cantidad suficiente de datos sintéticos y probablemente mejores datos sintéticos como los producidos por una CVAE-GAN, es posible que se pueda igualar o hasta mejorar el desempeño predictivo que usando solo datos reales, ya que **se usan solo datos reales para la validación y test.**

12.4.2.2. ANÁLISIS DE R2 PARA CORTO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
transformer	12600	Serie de tiempo	0,999997	0,999918	0,999956	0,00%	0,00%
transformer	25200	Serie de tiempo	0,999998	0,999937	0,999959	0,00%	0,00%
transformer	50400	Serie de tiempo	0,999998	0,999950	0,999968	0,00%	0,00%

Tabla 42 - Resultados R2 de experimento de uso de datos sintéticos para corto plazo

La Tabla 42 muestra el resultado del coeficiente de determinación (R^2) para el modelo Transformer en predicciones a corto plazo revela un rendimiento prácticamente perfecto y extremadamente consistente en los conjuntos de entrenamiento, prueba y validación, con valores que se acercan a 1 en todas las configuraciones de cantidad de datos. Esto indica que el modelo explica una proporción casi total de la varianza en la variable objetivo para las tres particiones. La ausencia de variación entre el rendimiento del entrenamiento y la validación, así como entre el de prueba y validación, sugiere un ajuste ideal sin evidencia de sobreajuste o subajuste, demostrando una robustez y capacidad de generalización sobresalientes a datos no vistos para esta métrica.

12.4.2.3. ANÁLISIS DE SNR PARA CORTO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
transformer	12600	Serie de tiempo	6.188,34	6.950,87	5.384,28	-12,99%	-22,54%
transformer	25200	Serie de tiempo	8.409,41	7.411,13	5.611,29	-33,27%	-24,29%
transformer	50400	Serie de tiempo	10.488,64	6.927,94	5.255,90	-49,89%	-24,13%

Tabla 43 - Resultados SNR de experimento de uso de datos sintéticos para corto plazo

La Tabla 43 muestra los resultados de SNR para el modelo Transformer en predicciones a corto plazo revela una tendencia donde el SNR en el conjunto de entrenamiento es significativamente más alto que en los conjuntos de validación y prueba, con una diferencia que se acentúa a medida

que aumenta la cantidad de datos. Esta "Variación Train - Validación" consistentemente negativa y creciente sugiere un sobreajuste del modelo a los datos de entrenamiento, donde está capturando ruido o patrones específicos que no se generalizan a datos no vistos. Aunque el SNR en validación y prueba se mantiene en un rango más bajo, la "Variación Test - Validación" es consistentemente negativa y considerable, lo que indica un rendimiento del modelo en el conjunto de prueba sistemáticamente inferior al observado en el conjunto de validación, lo que plantea preocupaciones sobre la robustez y capacidad de generalización real del modelo en un escenario de trading en vivo para esta métrica.

12.4.2.4. ANÁLISIS DE MAE PARA LARGO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
cnn	12600	Serie de tiempo	0,004807	0,004420	0,005194	8,05%	17,51%
cnn	25200	Serie de tiempo	0,005022	0,003909	0,005293	5,40%	35,41%
cnn	50400	Serie de tiempo	0,003491	0,004612	0,006598	89,00%	43,06%

Tabla 44 - Resultados MAE de experimento de uso de datos sintéticos para largo plazo

La Tabla 44 muestra para el modelo CNN en predicciones a largo plazo revela valores generalmente bajos en los conjuntos de entrenamiento, prueba y validación. Se observa una tendencia del MAE de entrenamiento a fluctuar, mientras que en validación y prueba se mantiene en un rango bajo. La "Variación Train - Validación" muestra un incremento significativo con la mayor cantidad de datos, lo que podría indicar un posible subajuste o dificultad del modelo para aprender de la complejidad adicional de los conjuntos de datos más grandes. Adicionalmente, la "Variación Test - Validación" es consistentemente alta y aumenta con la cantidad de datos, sugiriendo una menor capacidad de generalización del modelo a datos no vistos para esta métrica a largo plazo, siendo más sensible a las particularidades del conjunto de prueba.

12.4.2.5. ANÁLISIS DE R2 PARA LARGO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
cnn	12600	Serie de tiempo	0,996068	0,837669	0,945701	-5,06%	12,90%
cnn	25200	Serie de tiempo	0,996082	0,869821	0,943690	-5,26%	8,49%
cnn	50400	Serie de tiempo	0,992531	0,810620	0,904434	-8,88%	11,57%

Tabla 45 - Resultados R2 de experimento de uso de datos sintéticos para largo plazo

La Tabla 45 muestra los resultados para el modelo CNN en predicciones a largo plazo revela un rendimiento generalmente alto en el conjunto de entrenamiento, con valores cercanos a 1, lo que indica una excelente capacidad para explicar la varianza en los datos de entrenamiento. Sin embargo, se observa una disminución gradual del R^2 en los conjuntos de prueba y validación a medida que aumenta la cantidad de datos, lo que sugiere una posible reducción en la capacidad de generalización del modelo a medida que se expone a más complejidad. La "Variación Train - Validación" es consistentemente baja y negativa, lo que indica un sobreajuste limitado, mientras que la "Variación Test - Validación" es consistentemente positiva y relativamente alta, lo que podría señalar una mayor sensibilidad del modelo a las características específicas del conjunto de prueba.

12.4.2.6. ANÁLISIS DE SNR PARA LARGO PLAZO

Modelo	Cantidad de datos	Tipos de datos	Train	Test	Validación	Variación Train - Validación	Variación Test - Validación
cnn	12600	Serie de tiempo	907,36	1.059,15	956,86	5,46%	-9,66%
cnn	25200	Serie de tiempo	849,97	1.043,89	887,60	4,43%	-14,97%
cnn	50400	Serie de tiempo	12.546,20	12.730,21	13.726,11	9,40%	7,82%

Tabla 46 - Resultados SNR de experimento de uso de datos sintéticos para corto plazo

La Tabla 46 deja ver el SNR para el modelo CNN en predicciones a largo plazo revela un comportamiento variable. En las cantidades de datos más pequeñas, el SNR en entrenamiento es ligeramente inferior al de validación y prueba, con una variación positiva entre entrenamiento y validación que sugiere un ligero sobreajuste inicial. Sin embargo, con 50400 datos, el SNR en entrenamiento y prueba aumenta drásticamente y supera al de validación. El salto significativo en el SNR para 50400 datos en todos los conjuntos, especialmente en entrenamiento y prueba, podría indicar un cambio en el aprendizaje o una captura más efectiva de la señal en ese volumen de datos, o bien, una mayor sensibilidad a la estructura de la serie temporal.

13. ANÁLISIS DE RESULTADOS DE SIMULACIÓN DE TRADING

Esta sección presenta un análisis de los resultados obtenidos de la simulación de la **estrategia heurística de trading descrita en la sección 6.7** para los diferentes experimentos del proyecto de predicción de la tasa de cambio EUR/USD. Las simulaciones se realizaron utilizando esta estrategia heurística que combina predicciones de corto plazo (corto plazo, horaria) y largo plazo (largo plazo, diaria o 24 horas).

El objetivo es extraer *insights* sobre cómo diferentes configuraciones de datos y modelos impactan el rendimiento en un entorno de trading simulado, considerando métricas como *Profit*, número de operaciones (*N. Trades*), porcentaje de operaciones ganadoras (*% Ganadoras*), máxima reducción de capital después de un máximo (*Max Drawdown*), *Ratio de Sharpe* y *Riesgo*.

Parámetros básicos de simulación aplicables a todos los experimentos analizados:

- **Balance Inicial:** USD 10,000.
- **Apalancamiento** (Leverage): 100, determina la máxima cantidad de lotes que puede comprar con el Tamaño de Orden (porcentaje del balance a usar) donde se pueden comprar/vender hasta $Tamaño\ de\ Lote = Leverage * Tamaño\ de\ Orden / 100k$.
- **Tamaño de Lote:** un lote equivale a 100.000 de unidades de moneda, el mínimo lote posible a comprar/vender con apalancamiento, es de 1000 unidades de moneda (0.1 Lotes) y máximo 1.000.000 unidades de moneda (10 lotes).
- **Tamaño del Orden:** máximo 3% del balance por operación. Este valor en la estrategia se calcula proporcionalmente al beneficio/riesgo entre el mínimo tamaño y 3%.
- **Tasa Libre de Riesgo para Ratio de Sharpe:** Se utilizó un 10% anual, basada tasas aproximadas de instrumentos disponibles en USA/Europa.
- Se configuraron las siguientes **Tarifas** para una simulación más realista con valores típicos cobrados por Brokers de Forex reales:
 - **Spread:** Es la diferencia entre el precio de compra (ask) y de venta (bid) expresada en pips; se cobra sobre el volumen de la operación (aprox 0.02%)
 - **Slippage:** Es la variación entre el precio esperado y el precio real de ejecución; se aplica como un coste adicional sobre el valor de la operación (0.005%)
 - **Swap for long** positions per year: -2.2%, es el coste de dejar una posición abierta de compra durante un año.
 - **Swap for short** positions per year: -2.2%, es el coste de dejar una posición abierta de venta durante un año.

13.1. PARÁMETROS OPTIMIZADOS DE LA ESTRATEGIA DE TRADING HEURÍSTICA

Aparte de los parámetros básicos, la estrategia de trading heurística utilizada en todas las simulaciones se configuró con los siguientes cinco parámetros optimizables que a diferencia de los

parámetros básicos(fijos), son optimizables. Estos parámetros fueron optimizados utilizando la librería DEAP de algoritmos genéticos para la mejor predicción de largo plazo y corto plazo obtenida en todos los experimentos (sección 14) y luego, estos parámetros optimizables se mantuvieron fijos para todos los experimentos presentados en este análisis:

- **profit_threshold:** Especificado en pips, representa la ganancia mínima proyectada por las predicciones que una orden debe tener para ser considerada. Se evalúa en el momento de decidir si se abre una nueva operación.
- **tp_multiplier:** Es un factor multiplicador que se aplica al nivel de *Take Profit* óptimo predicho por los modelos. Su propósito es introducir un margen de tolerancia ante posibles errores en la predicción exacta del punto de toma de ganancias.
- **sl_multiplier:** Similar al anterior, este es un factor multiplicador para el nivel de *Stop Loss* óptimo predicho. Permite ajustar la tolerancia al riesgo ante errores en la predicción del punto de detención de pérdidas.
- **lower_rr_threshold:** Este umbral inferior para la relación recompensa/riesgo (*reward/risk ratio* o RR) influye en la decisión del tamaño de la orden. El RR se calcula como el *Profit* potencial dividido por el *Drawdown (caída después de máximo)* potencial entendido como la distancia en pips hasta el pico inverso después de un máximo o mínimo precio en las predicciones. Si el RR calculado para una nueva operación (ya sea de compra o venta) está por debajo de este `lower_rr_threshold`, se utiliza el mínimo tamaño de orden permitido por la estrategia.
- **upper_rr_threshold:** Si el RR calculado para una nueva operación supera este umbral superior, se utiliza el máximo tamaño de orden permitido por la estrategia (configurado como el 3% del balance total de la cuenta en el momento de la operación).

Estos parámetros buscan equilibrar la búsqueda de oportunidades rentables con una gestión de riesgo prudente, adaptando la exposición en cada operación según la calidad percibida de la señal de trading derivada de las predicciones.

13.2. VARIACIÓN DE CANTIDAD DE REGISTROS DE ENTRENAMIENTO

En este experimento, se evaluó el impacto de variar la cantidad de registros de entrenamiento (de 1575 a 25,200) en los modelos base (ANN, CNN, LSTM, Transformer). Para cada modelo y cantidad de datos específica, dicho modelo generó tanto las predicciones horarias (corto plazo) como las de 24 horas (largo plazo) utilizadas por la estrategia de simulación. Los modelos fueron entrenados utilizando las características base (OHLC y 8 derivadas simples) junto con las 11 características provenientes de la descomposición de series de tiempo.

Nótese que el Riesgo en este caso se calculó como la máxima caída posterior al máximo balance. Pero se observó que esta métrica no es representativa del riesgo por el conjunto de trades como lo es el Sharpe Ratio y puede ser desorientador, así que se optó por no tener en cuenta esta métrica.

Modelo Bajo Prueba	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	1575	-214.00	13	23.08	2.36	-54.38
ANN	3150	324.17	23	39.13	2.48	-7.27
ANN	6300	409.16	40	35.00	3.55	-6.29
ANN	12600	-178.19	37	27.03	6.05	-16.46
ANN	25200	115.62	31	35.48	3.37	-11.52
CNN	1575	152.13	13	38.46	2.23	-8.78
CNN	3150	-755.80	34	29.41	8.22	-30.68
CNN	6300	185.42	12	33.33	2.05	-5.59
CNN	12600	-332.14	12	16.67	3.78	-61.60
CNN	25200	193.75	20	35.00	3.88	-7.77
LSTM	1575	-101.58	1	0.00	1.47	0.00
LSTM	3150	630.77	20	40.00	2.35	2.74
LSTM	6300	-235.36	32	21.88	4.38	-13.06
LSTM	12600	322.65	19	36.84	3.02	-6.27
LSTM	25200	-102.50	2	0.00	1.48	-217.42
Transformer	1575	-263.42	40	22.50	7.76	-12.49
Transformer	3150	-669.03	21	14.29	7.37	-72.79
Transformer	6300	-560.67	17	11.76	6.31	-70.62
Transformer	12600	607.91	21	38.10	2.00	3.09
Transformer	25200	-111.97	27	25.93	2.54	-13.91

Tabla 47 - Resultados de Simulación (Baja Frecuencia) - Experimento: Variación de Cantidad de Registros

Modelo Bajo Prueba	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	1575	-202.85	17	29.41	2.43	-57.05
ANN	3150	-104.42	28	14.29	4.94	-10.76
ANN	6300	-76.45	51	23.53	8.84	-12.95
ANN	12600	9.86	52	23.08	7.87	-10.44
ANN	25200	451.91	39	35.90	2.89	-6.05
CNN	1575	-14.05	17	29.41	3.83	-11.70
CNN	3150	-620.97	49	26.53	12.07	-18.54
CNN	6300	452.81	14	35.71	2.05	-3.57
CNN	12600	-142.89	13	23.08	2.62	-19.08
CNN	25200	111.15	21	33.33	4.66	-8.65
LSTM	1575	-101.58	1	0.00	1.47	0.00
LSTM	3150	83.70	25	20.00	5.28	-8.07
LSTM	6300	153.96	41	19.51	5.89	-7.06
LSTM	12600	287.26	32	25.00	5.62	-6.93
LSTM	25200	-102.50	2	0.00	1.48	-217.42
Transformer	1575	104.68	47	27.66	7.00	8.32
Transformer	3150	-693.69	29	17.24	7.98	-39.53
Transformer	6300	-406.33	19	21.05	5.70	-30.42
Transformer	12600	129.94	25	20.00	4.94	-7.75
Transformer	25200	-270.79	32	15.62	5.60	-13.53

Tabla 48 - Resultados de Simulación (Alta Frecuencia) - Experimento: Variación de Cantidad de Registros

Al analizar los resultados de este experimento para los modelos base (ANN, CNN, LSTM, Transformer), se observa una dinámica compleja. El impacto de la cantidad de datos en el *Profit* y el *Ratio de Sharpe* parece depender intrínsecamente del tipo de modelo y la frecuencia de la operativa, sin una tendencia lineal clara y universal para todos los modelos.

Análisis de resultados en en Baja Frecuencia (BF) el desempeño de trading fue el siguiente:

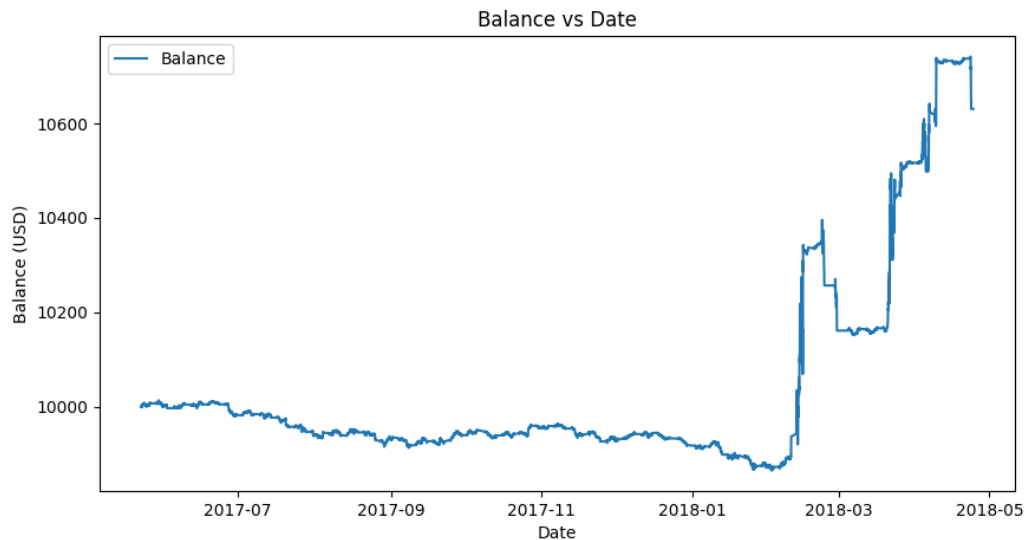


Figura 32 - Balance de Trading usando predicciones con LSTM entrenado con 3150 registros.

Los modelos **ANN** muestran un comportamiento no monotónico; el *Profit* es positivo con 3150 (\$324.17) y 6300 (\$409.16) registros pero negativo con 12600 y 1575, y modestamente positivo con 25200 (\$115.62). Los *Ratios de Sharpe* son mayormente negativos. Los modelos **CNN** también tienen un rendimiento variable; el *Profit* más alto se da con 6300 registros (\$185.42), pero con 3150 y 12600 registros el resultado es negativo. Los modelos **LSTM** muestran el mejor *Profit* (\$630.77) y el único *Ratio de Sharpe* positivo significativo (2.74) con 3150 registros, estos resultados se muestran en la **Figura 33**. Con otras cantidades de datos, el rendimiento es pobre o negativo. Los modelos **Transformer** en BF generalmente muestran *Profits* negativos, con una excepción para 12600 registros donde el *Profit* es de \$607.91 y el *Ratio de Sharpe* es positivo (3.09).

En Alta Frecuencia (AF) el desempeño de trading fue el siguiente:

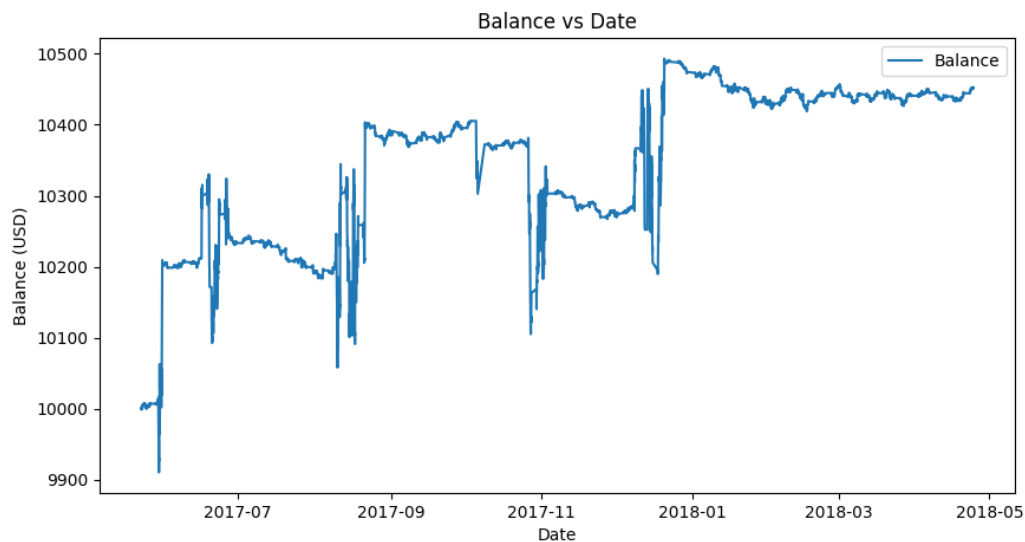


Figura 33 - Balance de Trading usando predicciones con LSTM entrenado con 12600 registros.

El modelo **ANN** con 25,200 registros presenta el *Profit* más alto en esta categoría (\$451.91), aunque su *Ratio de Sharpe* es negativo (-6.05). Con menos datos, los *Profits* son negativos o muy bajos. Los modelos **CNN** en AF tienen su mejor *Profit* con 6300 registros (\$452.81), pero los *Ratios de Sharpe* son consistentemente negativos. Los modelos **LSTM** en AF muestran *Profits* modestos y *Ratios de Sharpe* mayormente negativos. El mejor *Profit* (\$287.26) es con 12600 datos, este resultado se muestra en la **Figura 34**. Los modelos **Transformer** en AF tienen un comportamiento mixto, con el mejor *Profit* (\$129.94) para 12600 registros, y un *Ratio de Sharpe* positivo (8.32) para 1575 registros, aunque este último con *Profit* modesto.

El número de operaciones (*N. Trades*) no sigue un patrón simple con el aumento de datos. El porcentaje de operaciones ganadoras (*% Ganadoras*) se mantiene generalmente entre el 10% y el 40%. Un *Profit* positivo con un bajo porcentaje de ganadoras (e.g., LSTM 3150 BF, Transformer 12600 BF) sugiere que las operaciones ganadoras fueron significativamente más rentables que las perdedoras.

Con base en la información, se puede llegar a concluir que la relación entre la cantidad de datos de entrenamiento y el rendimiento en trading es compleja y no uniforme entre los diferentes modelos y frecuencias de operación. No se observa una mejora consistente en todas las métricas simplemente por aumentar la cantidad de datos. Algunos modelos y configuraciones específicas (ej. LSTM con 3150 registros en BF, Transformer con 12600 en BF, y ANN o CNN con cantidades de datos medias a altas en AF para *Profit* bruto) muestran picos de rendimiento, pero la mayoría de las configuraciones presentan *Ratios de Sharpe* negativos, indicando que, bajo esta estrategia, los modelos base luchan por generar una rentabilidad que supere consistentemente la tasa libre de riesgo asumida, ajustada por su volatilidad. La alta variabilidad en los resultados subraya la dificultad de la tarea de predicción y la sensibilidad de la estrategia de trading a la calidad de las predicciones.

13.3. VARIACIÓN DEL TIPO DE CARACTERÍSTICAS

En este experimento, se evalúa el impacto de añadir diferentes tipos de características a los modelos. Todos los modelos se entrenan con 25.200 registros y el conjunto específico de características adicionales de cada sub-experimento. Cada modelo/configuración genera ambas predicciones (corto plazo y largo plazo) para la simulación.

13.3.1. ADICIÓN DE INDICADORES TÉCNICOS (12 IT)

Se añadió un conjunto de 12 indicadores técnicos a los modelos (ANN, CNN, LSTM, Transformer), todos entrenados con 25,200 registro

Modelo Bajo Prueba (+IT)	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	-346.13	36	22.22	6.70	-17.53
CNN	25200	-634.64	35	22.86	7.02	-33.75
LSTM	25200	-314.53	21	19.05	3.37	-22.96
Transformer	25200	-449.69	34	23.53	6.83	-22.41

Tabla 49 - Resultados de Simulación (Baja Frecuencia) - Adición de Indicadores Técnicos

Modelo Bajo Prueba (+IT)	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	814.61	50	24.00	5.93	-1.55
CNN	25200	-643.65	55	18.18	9.01	-20.69
LSTM	25200	-175.54	32	18.75	5.49	-14.74
Transformer	25200	-677.05	47	19.15	10.40	-23.15

Tabla 50 - Resultados de Simulación (Alta Frecuencia) - Adición de Indicadores Técnicos

Al incorporar 12 indicadores técnicos a los modelos (todos entrenados con 25,200 registros), los resultados de la simulación de trading son predominantemente negativos en términos de *Profit* y *Ratio de Sharpe*, tanto en baja como en alta frecuencia. La única excepción es el modelo ANN en alta frecuencia, que muestra un *Profit* de \$814.61, aunque su *Ratio de Sharpe* sigue siendo negativo (-1.55). El % *Ganadoras* se mantiene bajo, generalmente por debajo del 25%. Estos resultados sugieren que, bajo la configuración actual de la estrategia de trading, la simple adición de este conjunto de indicadores técnicos no se traduce en una mejora de la rentabilidad y, en muchos casos, la empeora en comparación con algunos de los escenarios del Experimento de Variación de Cantidad de Datos.

13.3.2. ADICIÓN DE DATOS ESTACIONALES (3 EST.)

Se añaden 3 características de estacionalidad (hora del día, día de la semana, día del mes) a los modelos (ANN, CNN, LSTM, Transformer), todos entrenados con 25.200 registros.

Modelo Bajo Prueba (+Est.)	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	-389.84	40	27.50	5.31	-21.68
CNN	25200	-204.86	14	14.29	3.56	-18.68
LSTM	25200	-389.23	27	22.22	6.22	-20.75
Transformer	25200	-295.18	8	12.50	3.21	-65.79

Tabla 51 - Resultados de Simulación (Baja Frecuencia) - Adición de Datos Estacionales

Modelo Bajo Prueba (+Est.)	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	401.40	45	24.44	3.51	-5.11
CNN	25200	-246.68	15	13.33	3.97	-20.06
LSTM	25200	-568.73	38	18.42	8.05	-24.22
Transformer	25200	289.52	44	22.73	6.00	-6.68

Tabla 52 - Resultados de Simulación (Alta Frecuencia) - Adición de Datos Estacionales

La adición de 3 características estacionales a los modelos, en general, no conduce a una mejora significativa en el rendimiento del trading; de hecho, la mayoría de las configuraciones resultan en *Profits* negativos y *Ratios de Sharpe* también negativos. Los modelos ANN y Transformer en alta frecuencia son los únicos que muestran un *Profit* positivo (\$401.40 y \$289.52 respectivamente), pero sus *Ratios de Sharpe* (-5.11 y -6.68) indican un rendimiento pobre ajustado al riesgo. El % *Ganadoras* permanece bajo, similar a la sub-experimento anterior. Esto sugiere que estas características estacionales, tal como se han incorporado, no están siendo explotadas eficazmente por la estrategia o los modelos para generar operaciones rentables de manera consistente.

13.3.3. ADICIÓN DE DATOS DE ALTA FRECUENCIA

Se añaden 16 características de alta frecuencia a los modelos (ANN, CNN, LSTM, Transformer), todos entrenados con 25,200 registros.

Modelo Bajo Prueba (+AFrec)	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	74.74	37	29.73	5.58	-9.90
CNN	25200	447.59	31	32.26	3.69	-5.28
LSTM	25200	-395.85	31	19.35	7.11	-17.21
Transformer	25200	-530.25	19	15.79	5.53	-79.71

Tabla 53 - Resultados de Simulación (Baja Frecuencia) - Adición de Datos de Alta Frecuencia

Modelo Bajo Prueba (+AFrec)	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	735.80	60	28.33	7.00	2.38
CNN	25200	-365.21	48	18.75	7.54	-16.00
LSTM	25200	-357.65	46	19.57	7.54	-16.24
Transformer	25200	-6.46	42	21.43	4.83	-10.32

Tabla 54 - Resultados de Simulación (Alta Frecuencia) - Adición de Datos de Alta Frecuencia

La incorporación de 16 características de alta frecuencia presenta resultados mixtos. En baja frecuencia, el modelo CNN es el único que muestra un *Profit* positivo significativo (\$447.59), aunque con un *Ratio de Sharpe* negativo. El modelo ANN en BF tiene un *Profit* positivo pero modesto. En alta frecuencia, el modelo ANN es el único que logra un *Profit* positivo (\$735.80) y, notablemente, un *Ratio de Sharpe* positivo (2.38), lo cual es el primer *Sharpe* positivo robusto observado en este experimento hasta ahora. Los demás modelos (CNN, LSTM, Transformer) muestran *Profits* negativos en alta frecuencia con estas características. El % *Ganadoras* sigue siendo bajo en general. El desempeño del ANN en AF con datos de alta frecuencia es un punto destacable y sugiere que estas características pueden ser particularmente útiles para este modelo en un régimen de trading más activo.

13.3.4. ADICIÓN DE INDICADORES FUNDAMENTALES (2 FUND.)

Se añaden 2 indicadores fundamentales (S&P 500, VIX, procesados a frecuencia horaria) a los modelos (ANN, CNN, LSTM, Transformer), todos entrenados con 25,200 registros.

Modelo Bajo Prueba (+Fund.)	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	-435.27	34	23.53	5.05	-25.20
CNN	25200	104.70	25	28.00	3.78	-11.62
LSTM	25200	-336.92	21	19.05	3.60	-24.12
Transformer	25200	-356.14	11	9.09	3.79	-73.63

Tabla 55 - Resultados de Simulación (Baja Frecuencia) - Adición de Indicadores Fundamentales

Modelo Bajo Prueba (+Fund.)	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	398.32	49	26.53	4.32	5.35
CNN	25200	264.47	41	19.51	6.07	-6.49
LSTM	25200	107.27	48	20.83	4.05	8.66
Transformer	25200	-237.00	13	15.38	3.46	-21.88

Tabla 56 - Resultados de Simulación (Alta Frecuencia) - Adición de Indicadores Fundamentales

La adición de 2 indicadores fundamentales a los modelos presenta un panorama mixto. En baja frecuencia, todos los modelos, excepto el CNN (que tuvo un *Profit* modesto de \$104.70), resultaron en *Profits* negativos. En alta frecuencia, los modelos ANN y LSTM muestran *Profits* positivos (\$398.32 y \$107.27 respectivamente), y además, ambos logran *Ratios de Sharpe* positivos (5.35 y 8.66), lo cual es un hallazgo interesante, especialmente el alto *Sharpe* del LSTM a pesar de un *Profit* menor. El modelo CNN en AF tuvo un *Profit* positivo pero un *Sharpe* negativo. El modelo Transformer tuvo un rendimiento pobre en ambas frecuencias con estos indicadores. El % *Ganadoras* sigue siendo bajo en general. Parece que los indicadores fundamentales, aunque en un conjunto reducido, pueden tener un impacto positivo selectivo, especialmente para ANN y LSTM en alta frecuencia, mejorando la rentabilidad ajustada al riesgo.

13.3.5. ADICIÓN DE TODAS LAS CARACTERÍSTICAS

Se añaden 40 características (combinación de todas las anteriores) a los modelos (ANN, CNN, LSTM, Transformer), todos entrenados con 25,200 registros.

Modelo Bajo Prueba (+Todas)	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	2873.95	38	26.32	7.79	6.38
CNN	25200	2740.22	69	30.43	9.48	6.91
LSTM	25200	87.97	62	16.13	8.02	-5.32
Transformer	25200	1711.44	38	23.68	7.83	2.80

Tabla 57 - Resultados de Simulación (Baja Frecuencia) - Adición de Todas las Características

Modelo Bajo Prueba (+Todas)	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	3287.20	66	37.88	7.27	15.06
CNN	25200	2763.37	70	32.86	6.72	12.25
LSTM	25200	3565.16	39	48.72	7.80	15.32
Transformer	25200	1754.18	92	25.00	7.23	5.83

Tabla 58 - Resultados de Simulación (Alta Frecuencia) - Adición de Todas las Características

La incorporación del conjunto completo de 40 características adicionales a los modelos (entrenados con 25,200 registros) muestra resultados notablemente positivos y superiores a los sub-experimentos anteriores con tipos de características evaluadas individualmente, e incluso a muchos escenarios del experimento en el que se varió la cantidad de registros de entrenamiento.

En Baja Frecuencia (BF) el desempeño de trading fue el siguiente:

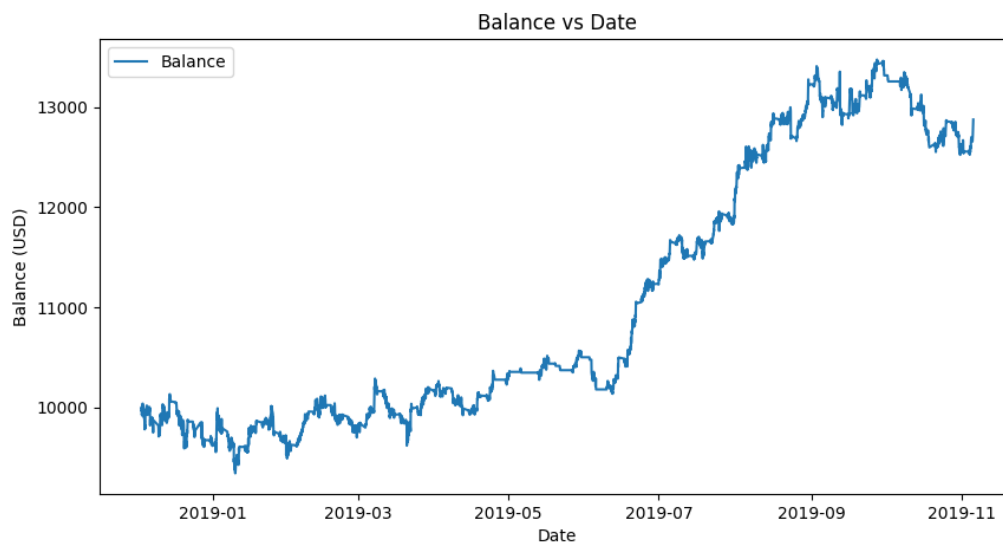


Figura 34 - Balance de Trading de baja frecuencia usando predicciones con ANN entrenado con 25200 registros.

Los modelos **ANN (Figura 35)** y **CNN** logran *Profits* significativos (\$2873.95 y \$2740.22 respectivamente) y, crucialmente, *Ratios de Sharpe* positivos y considerablemente altos (6.38 y 6.91). Esto indica una excelente rentabilidad ajustada al riesgo. El modelo **Transformer** también muestra un *Profit* robusto (\$1711.44) con un *Ratio de Sharpe* positivo (2.80). El modelo **LSTM** es la excepción en BF, con un *Profit* modesto y un *Ratio de Sharpe* negativo.

En Alta Frecuencia (AF) el desempeño de trading fue el siguiente:

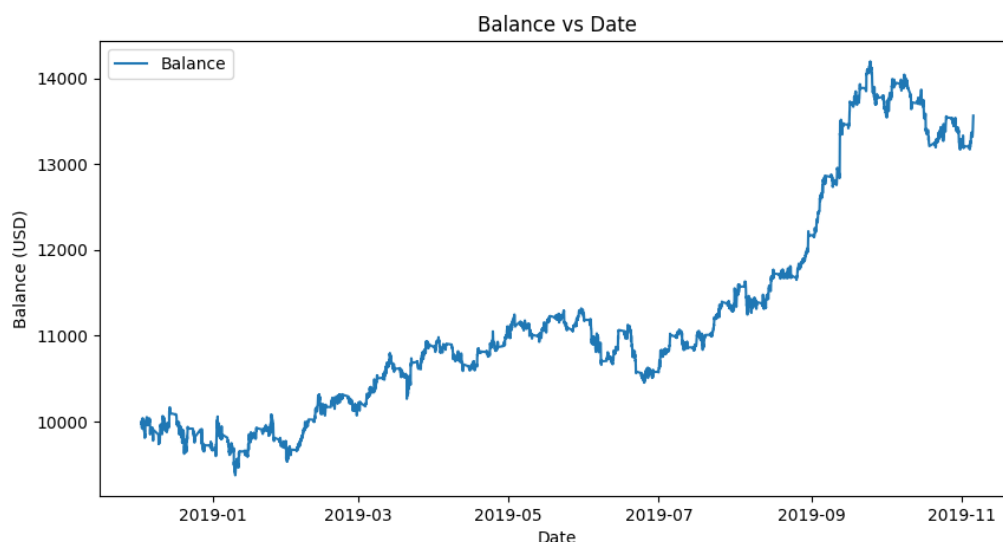


Figura 35 - Balance de Trading de alta frecuencia usando predicciones con LSTM entrenado con 25200 registros.

El rendimiento es aún más destacado. El modelo **LSTM** (Figura 36) tiene el *Profit* más alto (\$3565.16) y un excelente *Ratio de Sharpe* (15.32), además de un % *Ganadoras* del 48.72%, el más alto visto hasta ahora. Los modelos **ANN** y **CNN** también ofrecen *Profits* muy elevados (\$3287.20 y \$2763.37) con *Ratios de Sharpe* excepcionalmente buenos (15.06 y 12.25 respectivamente). El modelo **Transformer** en AF, aunque con el *Profit* más bajo del grupo (\$1754.18), también presenta un *Ratio de Sharpe* positivo y respetable (5.83).

Estos resultados sugieren fuertemente que un conjunto de características rico y diverso es altamente beneficioso para la estrategia de trading, permitiendo a los modelos capturar patrones más complejos y generar señales de trading más rentables y consistentes. El modelo LSTM, que tuvo un rendimiento mixto en experimentos anteriores, brilla particularmente en alta frecuencia cuando se le provee este conjunto completo de características.

Conclusiones generales para experimento de variación del tipo de características: La adición de diferentes tipos de características a los modelos base entrenados con 25,200 registros tiene un impacto variable. Mientras que la adición individual de Indicadores Técnicos, Datos Estacionales, o incluso Indicadores Fundamentales no mejoró consistentemente el rendimiento (y a menudo lo empeoró), la incorporación de Datos de Alta Frecuencia mostró promesa para el modelo ANN en alta frecuencia. Notablemente, la combinación de Todas las Características demuestra el impacto más significativamente positivo y robusto. Esta aproximación integral llevó a los *Profits* más altos y, crucialmente, a *Ratios de Sharpe* consistentemente positivos y elevados para la mayoría de los

modelos, tanto en baja como en alta frecuencia. Esto subraya el valor de un enfoque de ingeniería de características exhaustivo. El modelo LSTM, en particular, muestra un rendimiento excepcional en alta frecuencia con el conjunto completo de características, alcanzando el *Profit* y *Ratio de Sharpe* más altos de todo este experimento.

13.4. USO DE EXTRACTORES DE CARACTERÍSTICAS

En este experimento, se investiga el efecto de anteponer diferentes arquitecturas de Aprendizaje Profundo (ANN, CNN, LSTM, Transformer, cada una con dos capas), operando como extractores de características, a un **predictor fijo** (dos capas Conv1D -> capa LSTM -> capa Densa Bayesiana). Todos los modelos se entrenan con 25,200 registros. Las 12 características base (OHLC + 8 derivadas) alimentan al extractor, y su salida se combina con las 11 características de descomposición antes de ingresar al predictor fijo. Cada combinación de "Extractor + Predictor Fijo" genera ambas predicciones (corto y largo plazo) para su simulación.

13.4.1. EXTRACTORES SIN PRE-ENTRENAMIENTO

Los extractores y el predictor fijo se entrenan juntos (end-to-end). En este sub-experimento, se presentan los resultados para las cuatro arquitecturas de extractor (ANN, CNN, LSTM, Transformer) combinadas con el predictor fijo.

Extractor	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	3076.01	52	40.38	8.64	13.79
CNN	25200	2139.81	49	36.73	11.12	7.23
LSTM	25200	3248.52	52	40.38	8.75	14.79
Transf	25200	3226.27	63	39.68	8.37	15.50

Tabla 59 - Resultados de Simulación (Baja Frecuencia) - Extractores sin Pre-entrenamiento + Predictor Fijo

Extractor	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	3043.59	67	35.82	7.27	13.58
CNN	25200	2766.11	59	37.29	8.15	10.80
LSTM	25200	2779.48	64	34.38	7.22	12.14
Transf	25200	2967.68	76	32.89	7.27	13.34

Tabla 60 - Resultados de Simulación (Alta Frecuencia) - Extractores sin Pre-entrenamiento + Predictor Fijo

El uso de una arquitectura con un Extractor (ANN, CNN, LSTM, o Transformer) seguido de un Predictor Fijo muestra resultados consistentemente positivos y muy robustos. En baja frecuencia, los extractores LSTM y Transformer, combinados con el predictor fijo, permiten alcanzar los Ratios de Sharpe más altos (14.79 y 15.50 respectivamente). En alta frecuencia, el extractor ANN con el predictor fijo lidera en Ratio de Sharpe (13.58).

13.4.2. EXTRACTORES CON PRE-ENTRENAMIENTO (AUTOENCODER)

Los módulos extractores (ANN, CNN, LSTM, Transformer) son pre-entrenados usando un autoencoder antes de ser combinados con el predictor fijo.

Extractor	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	2612.75	48	37.50	6.72	10.91
CNN	25200	4298.06	52	48.08	8.85	20.21
LSTM	25200	3004.87	58	36.21	8.65	13.33
Transf	25200	2702.57	68	35.29	8.65	12.70

Tabla 61 - Resultados de Simulación (Baja Frecuencia) - Extractores Pre-entrenados + Predictor Fijo

Extractor	Registros	Profit (\$)	N. Trades	% Ganadoras	Riesgo (%)	Ratio de Sharpe
ANN	25200	3167.20	63	36.51	7.22	14.07
CNN	25200	5020.06	73	41.10	7.27	22.70
LSTM	25200	2916.53	79	30.38	7.23	12.98
Transf	25200	2796.51	97	30.93	7.43	13.25

Tabla 62 - Resultados de Simulación (Alta Frecuencia) - Extractores Pre-entrenados + Predictor Fijo

Interpretación de Resultados con Extractores Pre-entrenados:

El pre-entrenamiento de los extractores mediante un autoencoder mejora significativamente el rendimiento respecto a usar un extractor no pre-entrenado.

En Baja Frecuencia (BF):

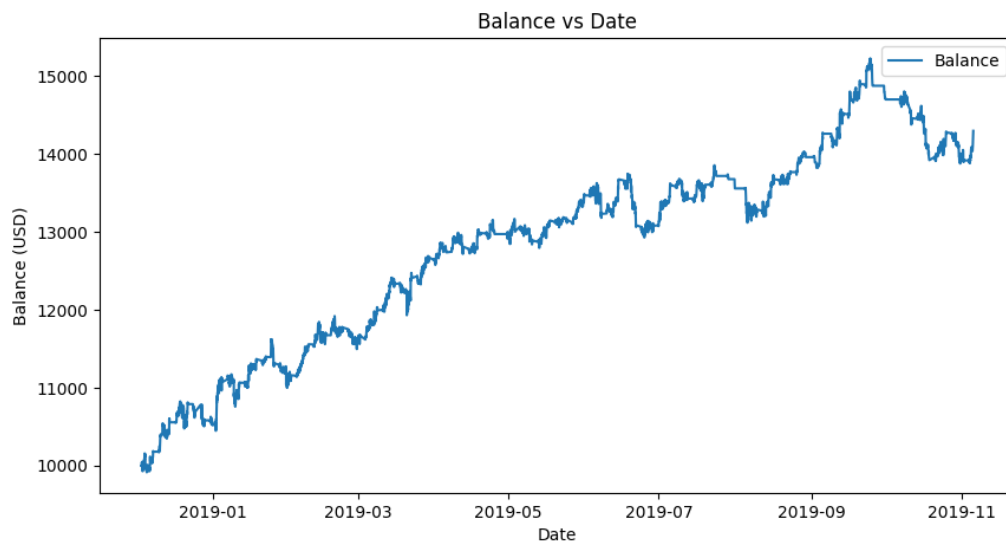


Figura 36 - Balance de Trading de Baja frecuencia usando predicciones con extractor CNN.

El Extractor CNN Pre-entrenado + Predictor Fijo es la combinación más destacada, logrando un Profit de \$4298.06 (Figura 37) y un Ratio de Sharpe sobresaliente de 20.21. El % Ganadoras también es muy alto (48.08%).

En Alta Frecuencia (AF):

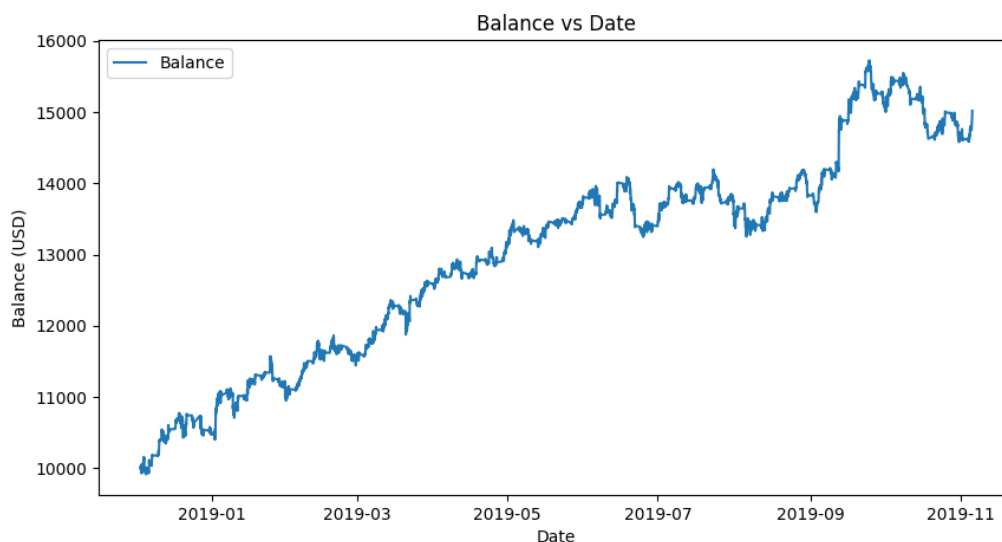


Figura 37 - Balance de Trading de Alta frecuencia usando predicciones con extractor CNN.

Nuevamente, el Extractor CNN Pre-entrenado lidera con un Profit de \$5020.06 y un Ratio de Sharpe de 22.70, el más alto observado en todos los experimentos. El % Ganadoras es del 41.10%. La Figura 38 muestra el balance plot para la simulación de trading usando las predicciones producidas por este modelo.

Estos resultados indican que el pre-entrenamiento del extractor CNN aporta un beneficio muy significativo al rendimiento de la estrategia de trading.

La arquitectura de dos etapas (Extractor + Predictor Fijo) es consistentemente beneficiosa. El pre-entrenamiento del extractor (especialmente el Extractor CNN Pre-entrenado) demuestra una mejora importante sobre los extractores sin pre-entrenamiento, llevando a los rendimientos más altos observados en todo el estudio.

Los datos sintéticos produjeron resultados decepcionantes en predicción como se observó en la sección 14 y los resultados fueron pésimos también en la simulación de trading.

13.5. RELACIÓN ENTRE PRECISIÓN DEL MODELO Y RENDIMIENTO DE TRADING

Para obtener una visión más holística de la relación entre la precisión de las predicciones de los modelos y su desempeño efectivo en la simulación de trading, se consolidaron los resultados de todos los experimentos. Específicamente, se tomó el *Profit* obtenido en cada simulación de trading y se contrastó con el Error Absoluto Medio (MAE) de la predicción de largo plazo (diaria) del

conjunto Test correspondiente al modelo y configuración de datos utilizados en dicha simulación. Adicionalmente, se analizó la relación entre el *Profit* y el *Ratio de Sharpe* obtenido en las simulaciones.

Estos análisis se presentan gráficamente mediante diagramas de dispersión (*scatter plots*), donde cada punto representa un experimento de simulación de trading. Se ha añadido una línea de tendencia exponencial para visualizar la relación general entre las variables.

13.5.1. RELACIÓN ENTRE PROFIT Y MAE DE TEST (LARGO PLAZO)

La Figura 38 muestra la relación entre el *Profit* de la simulación y el MAE de test de largo plazo para la operativa de **alta frecuencia**.

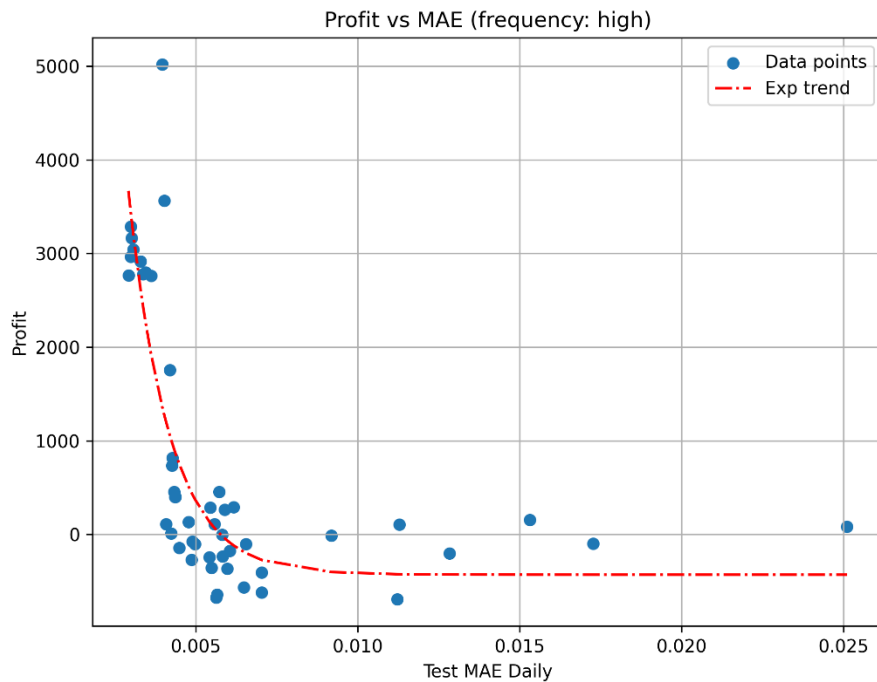


Figura 38 - Profit vs. MAE de Test (Largo Plazo) - Alta Frecuencia.

La Figura 39 presenta la misma relación para la operativa de **baja frecuencia**.

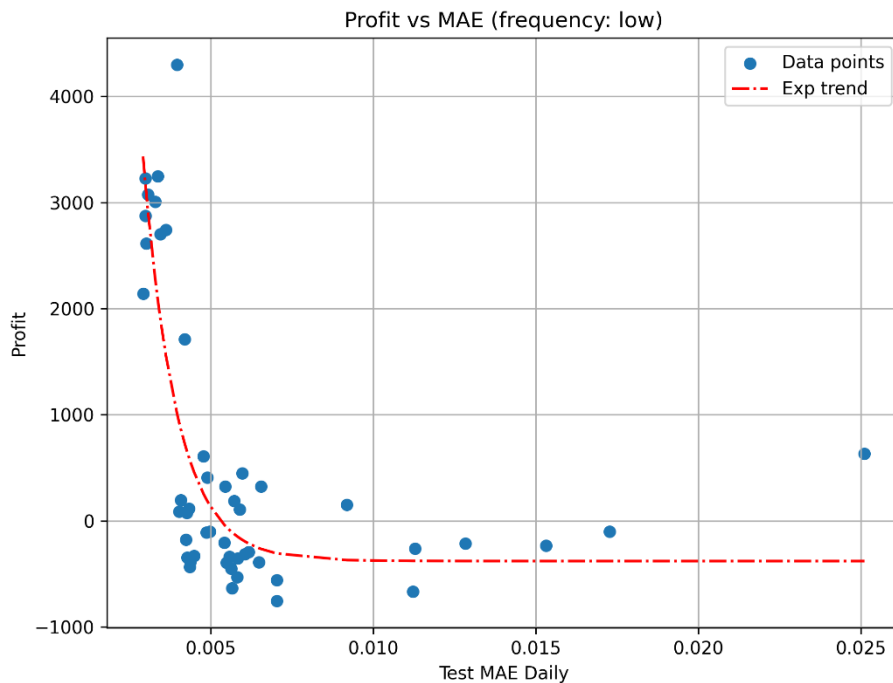


Figura 39 - Profit vs. MAE de Test (Largo Plazo) - Baja Frecuencia.

Interpretación de Profit vs. MAE: Observando ambos gráficos (Figura 38 y 39), se evidencia una tendencia general donde menores valores de MAE (es decir, predicciones de largo plazo más precisas) tienden a estar asociados con mayores Profits en la simulación de trading. La línea de tendencia exponencial sugiere que la relación no es lineal; pequeñas mejoras en la precisión (reducciones en el MAE), especialmente cuando el MAE ya es bajo, pueden llevar a incrementos desproporcionadamente mayores en el Profit. Esto es particularmente visible en los puntos que logran los Profits más altos, los cuales se concentran en la región de menor MAE. Sin embargo, también existe una dispersión considerable de puntos, lo que indica que el MAE no es el único factor determinante del Profit; otros aspectos como la calidad de la predicción de corto plazo, la robustez del modelo ante diferentes condiciones de mercado y la propia naturaleza de la estrategia de trading también juegan un papel crucial. La dispersión es más notable en los rangos de MAE más altos, donde los Profits pueden variar significativamente e incluso ser negativos.

Estos resultados se pueden contrastar con el resultado de la simulación de la estrategia de trading heurística usando **predicciones ideales de corto y largo plazo**, donde se alcanzan Profits de \$30k y \$35k respectivamente:

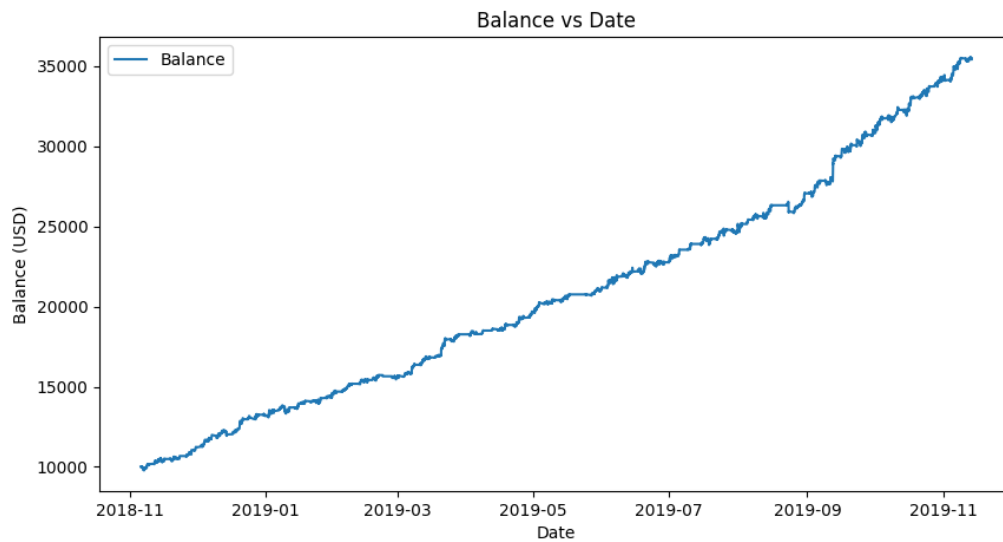


Figura 40 - Balance plot de trading con predicciones Ideales - Alta Frecuencia.

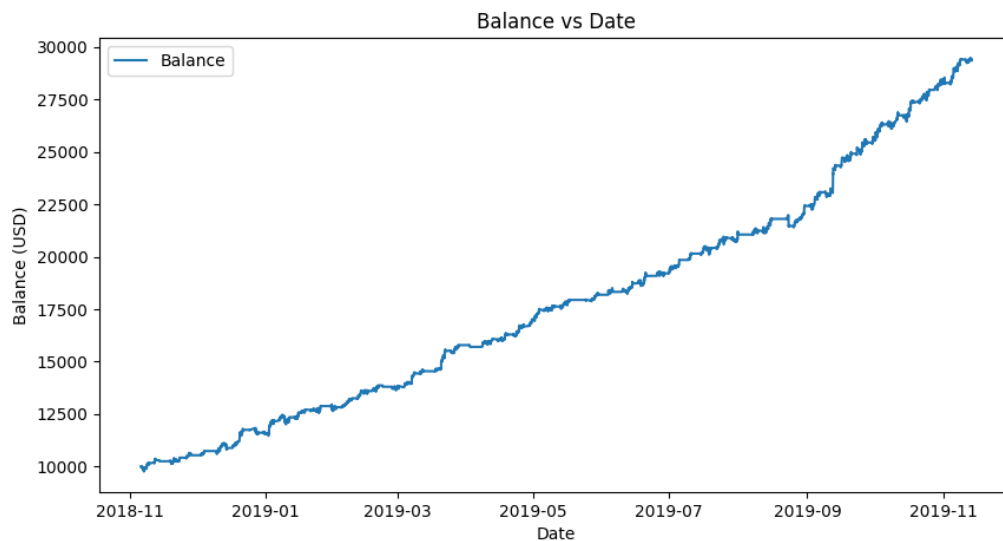


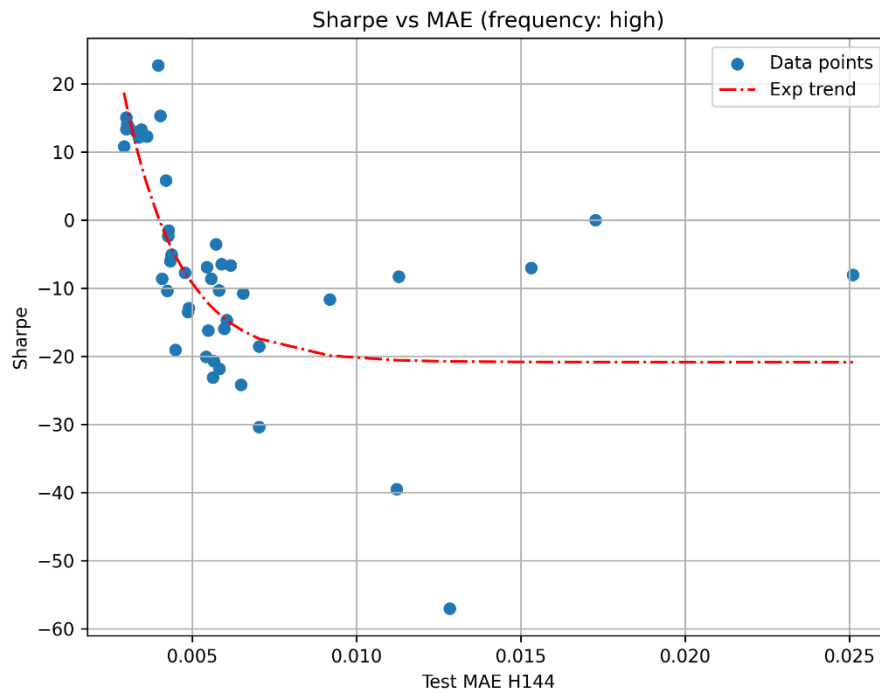
Figura 41 - Balance plot de trading con predicciones Ideales - Alta Frecuencia.

Como se observa en las Figuras 40 y 41, mientras con las mejores predicciones realizadas en este proyecto en alta frecuencia ($MAE=0.0004$) apenas se alcanza un profit de aproximadamente \$5k, con predicciones ideales ($MAE=0$) se obtiene aproximadamente \$35k. En baja frecuencia ocurre algo similar, mientras en la mejor predicción del proyecto se obtiene aprox \$4.2k, usando predicciones ideales se obtiene aproximadamente \$28k.

Estos resultados refuerzan la intuición de que pequeños decrementos en el error producen grandes incrementos en el Profit. También muestran que la estrategia heurística funciona casi perfectamente con datos ideales.

13.5.2. RELACIÓN ENTRE RATIO DE SHARPE Y MAE DE TEST (LARGO PLAZO)

La Figura 42 muestra la relación entre el *Ratio de Sharpe* de la simulación y el MAE de test de largo plazo para la operativa de **alta frecuencia**.



La Figura 43 presenta la misma relación para la operativa de baja frecuencia.

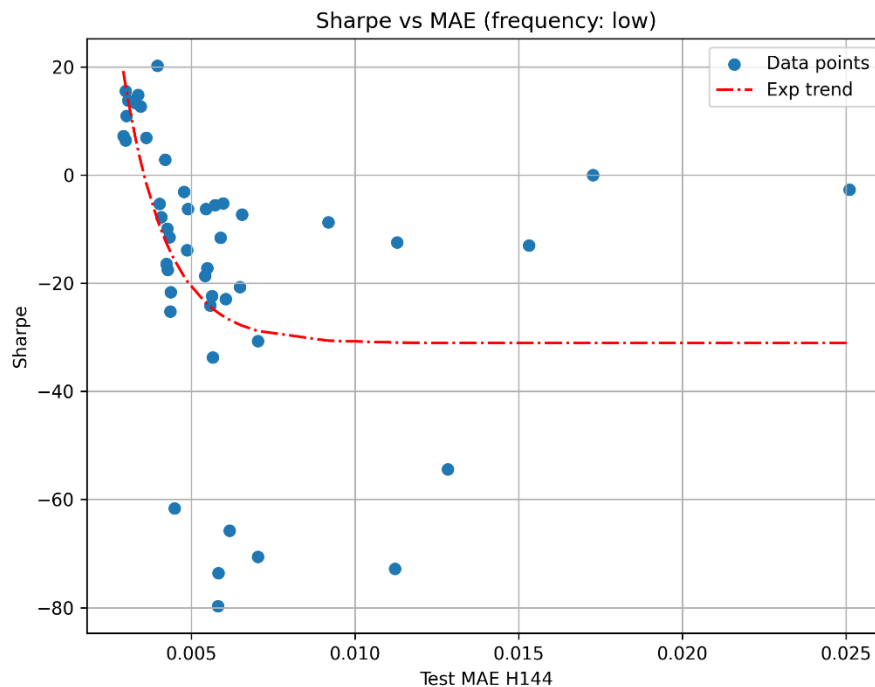


Figura 43 - Ratio de Sharpe vs. MAE de Test (Largo Plazo) - Baja Frecuencia.

Estos gráficos (Figura 42 y 43) exploran la relación entre la calidad de la predicción (inversamente proporcional al MAE) y la rentabilidad ajustada al riesgo (*Ratio de Sharpe*). Se esperaría idealmente que un menor MAE (mejor predicción) condujera a un mayor *Ratio de Sharpe*. En general, se observa una tendencia en esta dirección: los puntos con los *Ratios de Sharpe* más altos tienden a agruparse en la zona de menor MAE. La línea de tendencia exponencial sugiere que, a medida que el MAE disminuye, el potencial para alcanzar *Ratios de Sharpe* elevados aumenta de forma no lineal. Sin embargo, al igual que con el *Profit*, existe una dispersión considerable. Hay casos donde modelos con MAE relativamente bajos no logran *Ratios de Sharpe* altos (incluso pueden ser negativos), y viceversa, aunque esto último es menos común para *Ratios de Sharpe* muy elevados. Esto indica que si bien la precisión del modelo es un factor importante para obtener una buena rentabilidad ajustada al riesgo, no es el único. La consistencia de las predicciones, la capacidad del modelo para evitar errores grandes y la interacción con la dinámica de la estrategia de trading son también determinantes. Los *Ratios de Sharpe* muy negativos suelen estar asociados con un amplio rango de valores de MAE, lo que sugiere que un mal desempeño ajustado al riesgo puede ocurrir independientemente de la precisión absoluta del modelo si otros factores no son favorables.

14. CONCLUSIONES

Se evidenció que el incremento en la cantidad de registros, el incremento en las características adicionales usadas y el uso de pre-entrenamiento de extractores de características con un autoencoder producen mejoras en las predicciones que se traducen en mayores ganancias y menor riesgo en la simulación de Trading que usa estas predicciones.

Variación de cantidad de registros:

Se pudo observar un aumento en la cantidad de datos de entrenamiento en general conduce a una reducción del Error Absoluto Medio (MAE) y un aumento del coeficiente de determinación (R2) en los conjuntos de prueba y validación para todos los modelos (CNN, ANN, LSTM, Transformer). Esto sugiere que una mayor cantidad de datos permite a los modelos ajustarse para identificar patrones subyacentes con mayor precisión y mejorar su capacidad predictiva. Si bien la adición de más datos tiende a mejorar el rendimiento, la magnitud de esta mejora disminuye a medida que se disminuyen las métricas de Error, pero esos pequeños incrementos en la exactitud de las predicciones se traducen en ganancias cada vez mayores de la estrategia de trading basada en predicciones.

Este fenómeno de rendimientos decrecientes implica que existe un punto donde la incorporación de datos adicionales aporta beneficios marginales decrecientes en términos la reducción del error, pero estos beneficios marginales en la exactitud se traducen en ganancias cada vez más grandes en la estrategia de Trading, lo que hace que valga la pena realizar experimentos para mejorar aunque sea marginalmente las predicciones, ya que estas mejoras marginales en niveles muy bajos de error, se traducen en incrementos grandes en las ganancias en la estrategia de Trading como se pudo apreciar en los resultados.

Así mismo, se observan diferencias significativas en el rendimiento entre las arquitecturas de redes neuronales evaluadas para un mismo volumen de datos. En conjuntos de datos extensos, el modelo Transformer tiende a exhibir un MAE inferior y un R2 superior en comparación con otros modelos, lo que indica una mayor capacidad para predecir la tasa de cambio EURO/USD y explicar su varianza.

Uso de Características Adicionales:

El uso de características adicionales a las columnas base del precio (High, Low, Close, Open), como los indicadores técnicos, fundamentales, variables estacionales y de variables de alta frecuencia, produjeron un incremento en el desempeño de los predictores (menor MAE, mayor R2, mayor SNR, **Mayor profit en Trading**), siendo el mejor resultado usar todas las características adicionales juntas.

Uso de extractores de características pre-entrenados con Autoencoder:

Se observó una mejora del performance de prediccción y una importante **mejora en los resultados de trading** cuando se pre-entrena el extractor de características usando un autoencoder.

Los modelos predictivos con extractores de características pre-entrenados basados en CNN y Transformer muestran una mejor convergencia y capacidad de generalización, junto con los

mejores resultados de simulación de Trading, que los modelos predictivos que no usan el extractor de características pre-entrenado con el autoencoder.

El modelo extractor que usa Transformer, para el corto plazo, logra la mayor precisión (menor MAE). Todos los modelos alcanzan un ajuste bastante bueno (R2 muy alto). Para el largo plazo el extractor ANN, demuestra ser el más preciso (menor MAE) y con el mejor ajuste (R2 más alto) y menor incertidumbre (mayor SNR). Se observa una caída considerable en el R2 desde el entrenamiento para todos los modelos.

Datos Sintéticos:

La forma usada para generar datos sintéticos usando solo un VAE simple, no es suficiente para generar series de tiempo con calidad suficiente que puedan beneficiar el entrenamiento. Esto se debe a que el VAE genera series de tiempo episódicas, sin referencia a valores pasados y muchas veces se generan valores incoherentes, como por ejemplo en una fecha cuyo día de la semana el martes, el VAE puede generar esa señal estacional (día de la semana) como jueves o cualquier otro día. Por tanto, se hace necesario el uso de Conditional VAE, es decir que se entrene el autoencoder con entradas en el decoder para los datos como la fecha o los valores anteriormente generados para las características, de forma que le sirvan de entrada para producir series de tiempo condicionadas por estas entradas, reduciendo incoherencias en los datos generados. También se beneficia el generador de no generar los indicadores técnicos como con el VAE simple, sino a partir de una ventana de valores generados, calcula estos indicadores técnicos en lugar de generarlos.

Otras arquitecturas generativas que soportan entradas condicionales como GAN (CGAN) es casi seguro que darán resultados mejores que los generadores probados, pero por limitaciones de tiempo y capacidad computacional, no se implementaron.

Desempeño en Trading:

Se puede apreciar que **a medida que la exactitud de las predicciones aumenta, también aumenta la ganancia y mejora el Sharpe Ratio** durante la simulación de trading. Pero también se observó, sobre todo, que **incrementos muy pequeños en la exactitud de las predicciones (reducción de MAE), producen incrementos muy grandes en el Profit** de la estrategia de trading. Lo que hace pensar que el MAE en las predicciones y el Profit en esta estrategia particular tienen una relación que parece exponencial.

15. TRABAJO FUTURO

Entre los trabajos futuros más importantes está la creación de un cubo OLAP con migradores para los múltiples archivos de resultados producidos por los experimentos para facilitar el análisis de los resultados experimentales sin tediosas tareas de consolidación manual, pudiendo usar herramientas de BI para visualización y análisis más estandarizado de los resultados.

También se propone como trabajo futuro el uso de otros tipos de características con potencial de mejorar las predicciones, como uso como entradas de predicciones de modelos estadísticos como ARIMA o GARCH, también el uso de Clústering en las características de entrada para obtener un dataset de estados de mercado, o el uso de la técnica de Inferencia Causal para convertir el dataset de eventos de calendario económico en una serie de tiempo de influencia en la tendencia y volatilidad de precio para usar como entrada.

El uso de otros métodos de generación de datos sintéticos como GAN o VAE-GAN con entrada condicional, pueden beneficiar al entrenamiento y su implementación es un trabajo futuro prometedor.

La arquitectura Transformer demostró ser competitiva, especialmente en el horizonte a corto plazo con el extractor pre-entrenado. Futuras investigaciones podrían explorar la integración de mecanismos de atención temporal más sofisticados, como la atención dispersa (sparse attention) o la atención basada en kernel (kernel-based attention). Estas técnicas podrían permitir al modelo ponderar de forma más eficiente las partes relevantes de la secuencia temporal, mejorando así la capacidad de capturar patrones sutiles y dependencias no lineales, lo que podría traducirse en predicciones más precisas y robustas, especialmente en horizontes de predicción más extensos.

El tipo de cambio EURO/USD puede verse influenciado por eventos económicos, políticos y sociales, cuyo sentimiento puede reflejarse en fuentes de información no estructurada como noticias y redes sociales. Futuros trabajos podrían explorar la incorporación de técnicas de procesamiento del lenguaje natural (NLP) para extraer y cuantificar el sentimiento de estas fuentes de datos. Esta información de sentimiento podría luego integrarse como características adicionales en los modelos de redes neuronales. La hipótesis es que la inclusión de esta información cualitativa podría ayudar a los modelos a anticipar movimientos del tipo de cambio impulsados por el sentimiento del mercado, mejorando así la capacidad predictiva, especialmente en respuesta a eventos inesperados.

Así mismo el uso de Reinforcement Learning (NEAT) para Neuroevolucionar una red que controle los parámetros de las Estrategias de Trading de manera dinámica en lugar de usar parámetros estáticos para toda la simulación. Esta técnica consiste en un sistema de reinforcement learning para optimizar dinámicamente una red neuronal con 5 salidas para controlar los 5 parámetros de la estrategia de trading basada en las predicciones que actualmente se utilizan. El uso de un algoritmo como NEAT (NeuroEvolution of Augmenting Topologies) podría permitir la evolución de una red neuronal que controle los parámetros de la estrategia en tiempo real, adaptándose a las condiciones cambiantes del mercado en lugar de depender de parámetros estáticos.

16. REFERENCIAS BIBLIOGRÁFICAS

- [1] G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", 2003, *Neurocomputing*, Volume 50, pg.159-175
- [2] C. Brooks, "Introductory Econometrics for Finance", 2019, Cambridge University Press, 4th edition.
- [3] H. Kaijian et al, "Financial Time Series Forecasting with the Deep Learning Ensemble Model", 2023, *Mathematics*, Volume 11, Issue 4
- [4] Q. Hu, A. Yuille, Z. Zhou, "Synthetic Data as Validation", 2023, [online]. Available: <https://arxiv.org/abs/2310.16052>
- [5] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [6] L. Ștefan, B. Boteanu, C. Lamba, B. Ionescu, "Towards realistic financial time series generation via generative adversarial learning", 29th European Signal Processing Conference (EUSIPCO), 2021, pp. 1341–1345.
- [7] S. Siami-Namini, N. Tavakoli, A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series", 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 1394-1401, DOI: 10.1109/ICMLA.2018.00227.
- [8] M. Shabbeer, M. Pravalika, R. Pitchai, Ch. Madhubabu, "ARIMA - GRU Hybrid Model for Stock Exchange Forecast", 2023 9th International Conference on Smart Computing and Communications (ICSCC), 2023.
- [9] K. He, Q. Yang, "Financial Time Series Forecasting with the Deep Learning Ensemble Model", *Mathematics*, 2023, vol. 11, pp. 1054.
- [10] M. Buczyński, M. Chlebus, "Financial Time Series Models—Comprehensive Review of Deep Learning Approaches and Practical Recommendations", *Engineering Proceedings*, 2023, vol. 39, pp. 79.
- [11] R. Huang, X. Ren, "Research on the Application of Time Series Forecasting Model Based on Encoder-Decoder LSTM Model", 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), 2022
- [12] A. Zeng, M. Chen, L. Zhang, Q. Xu, "Are Transformers Effective for Time Series Forecasting?", *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 9, 2023.
- [13] F. Ramzan, C. Sartori, S. Consoli, D. Reforgiato Recupero, "Generative Adversarial Networks for Synthetic Data Generation in Finance: Evaluating Statistical Similarities and Quality Assessment", *AI*, vol. 5, no. 2, 2024
- [14] M. Dogariu, L.-D. Ștefan, B. A. Boteanu, C. Lamba, "Generation of Realistic Synthetic Financial Time-series", *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 18, no. 4, 2022

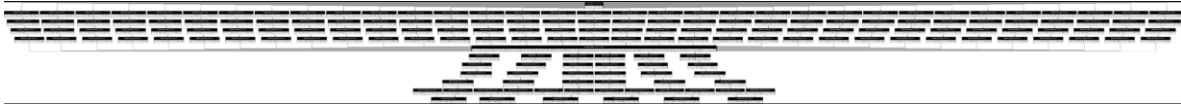
- [15] K. Miok, D. Nguyen-Doan, D. Zaharie, M. Robnik-Šikonja, "Generating Data using Monte Carlo Dropout", 2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP), 2019.
- [15] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge, U.K.: Cambridge University Press, 2003, Section 40, p. 483.
- [16] D. Caragea, "Unlocking the potential of bootstrapping: A journey towards balanced and reliable synthetic data. A framework for evaluating Bootstrap in the context of synthetic data generation", Master's Thesis, Utrecht University, 2024.
- [17] D. Malhotra, "Data Cleaning Techniques for Big Data and Machine Learning", Springer Briefs in Computer Science, 2023.
- [18] M. Hyndman, N. Athanassopoulos, "Forecasting: principles and practice", OTexts, 2023.
- [19] J. Shen, M. O. Shafiq, "Short-term stock market price trend prediction using a comprehensive deep learning system", *Journal of Big Data*, vol. 7, art. 66, 2020.
- [20] A. Ul Haq, A. Zeb, Z. Lei, D. Zhang, "Forecasting daily stock trend using multi-filter feature selection and deep learning", *Expert Systems with Applications*, vol. 168, 2021.
- [21] O. B. Sezer, M. U. Gudelek, A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019", *Applied Soft Computing*, vol. 90, 2020.
- [22] C. Bergmeir, M. Costantini, J. M. Benítez, "On the usefulness of cross-validation for directional forecast evaluation", *Computational Statistics & Data Analysis*, vol. 76, 2014
- [23] K. Chojnacki, R. Ślepaczuk, "Ensembled LSTM with Walk Forward Optimization in Algorithmic Trading", 2023
- [24] K. He, Q. Yang, L. Ji, J. Pan, Y. Zou, "Financial Time Series Forecasting with the Deep Learning Ensemble Model", *Mathematics*, vol. 11, no. 4, 2023.
- [25] H. B. C., "preprocessor," GitHub. Accessed: Oct. 13, 2024. [Online]. Available: <https://github.com/harveybc/preprocessor>.
- [26] S. M. Park, H. G. Yoon, D. B. Lee et al., "Topological magnetic structure generation using VAE-GAN hybrid model and discriminator-driven latent sampling," *Sci. Rep.*, vol. 13, p. 20377, 2023. doi: 10.1038/s41598-023-47866-3.
- [27] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, Jul./Oct. 1948.
- [28] S. Altuntaş, "Forecasting the Euro Exchange Rate Using Deep Learning: GRU vs. LSTM", *Journal of Risk and Financial Management*, vol. 16, no. 4, pp. 1–15, 2023.
- [29] L. Troiano y R. Restaino, "Foreign-exchange currency-rate prediction with Bidirectional-LSTM", *Proceedings*, vol. 68, no. 1, pp. 19–23, 2023.

[30] Q. Liang, Y. Chen, H. Zhang y S. Li, "Temporal Convolutional Networks for Financial Time Series Forecasting: A Survey", arXiv preprint arXiv:2303.12345, 2023.

ANEXOS

A continuación, se presentan las visualizaciones completas de las arquitecturas utilizadas en las fases experimentales. Por razones de espacio, se incluye una versión reducida en el documento y un enlace a la imagen de alta resolución en el repositorio del proyecto.

Anexo A. Arquitectura completa del modelo ANN



La imagen en alta resolución se encuentra disponible en:

https://github.com/harveybc/predictor/blob/master/examples/results/phase_2_1_daily/phase_2_1_ann_1d_model_plot.png

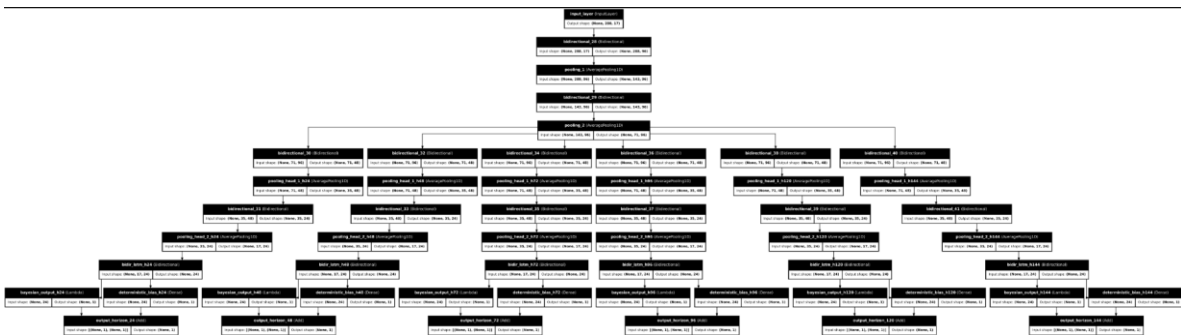
Anexo B. Arquitectura completa del modelo CNN



La imagen en alta resolución se encuentra disponible en:

https://github.com/harveybc/predictor/blob/master/examples/results/phase_2_2_daily/phase_2_2_cnn_1d_model_plot.png

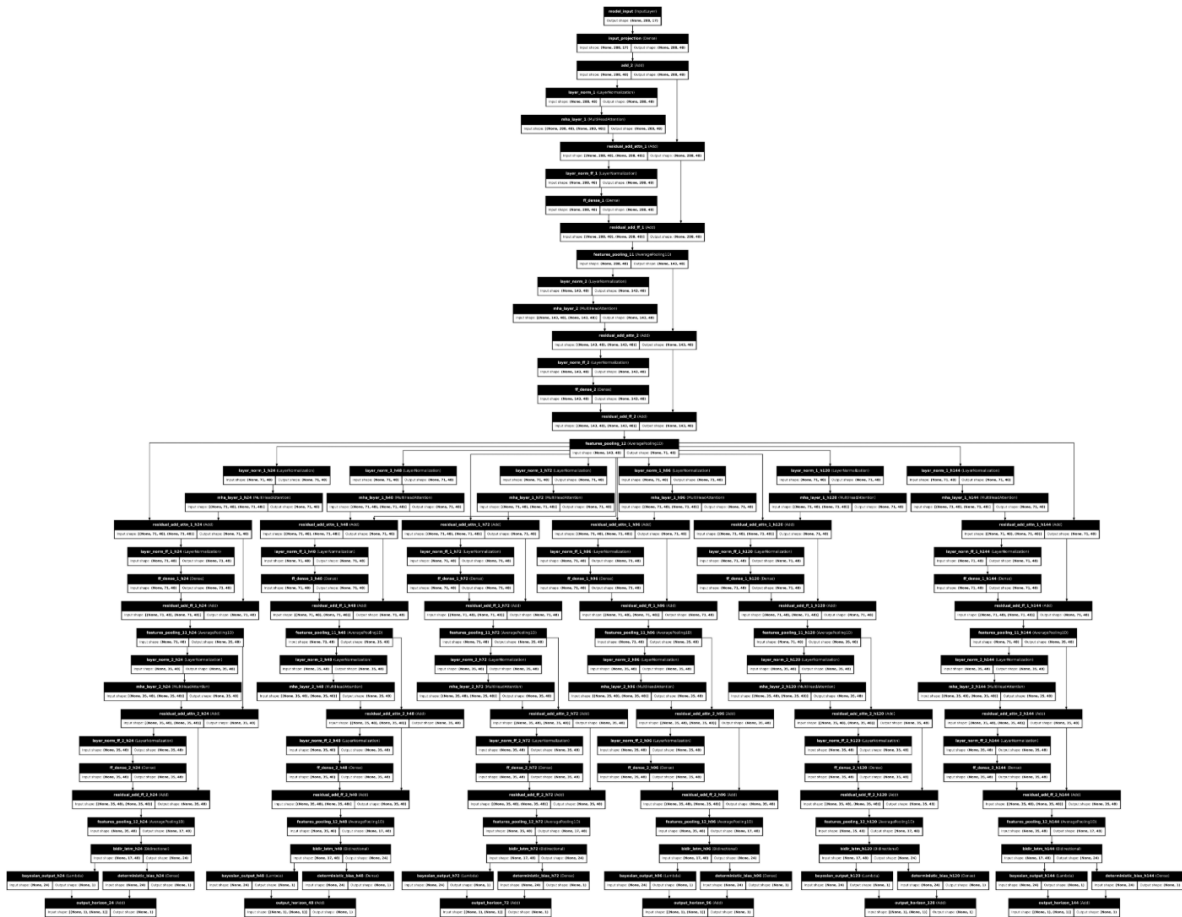
Anexo C. Arquitectura completa del modelo LSTM



La imagen en alta resolución se encuentra disponible en:

https://github.com/harveybc/predictor/blob/master/examples/results/phase_2_2_daily/phase_2_2_lstm_1d_model_plot.png

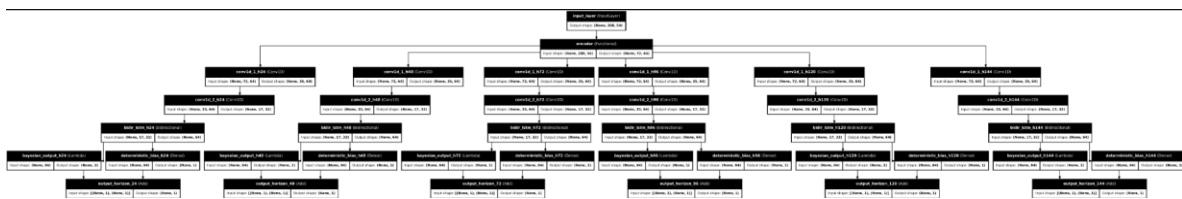
Anexo D. Arquitectura completa del modelo Transformers



La imagen en alta resolución se encuentra disponible en:

https://raw.githubusercontent.com/harveybc/predictor/refs/heads/master/examples/results/phase_2_2_daily/phase_2_2_transformer_1d_model_plot.png

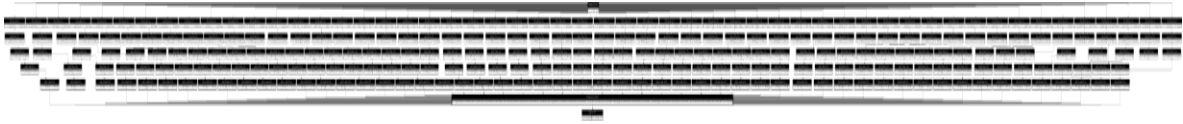
Anexo E. Cabezales predictivos autoencoder



La imagen en alta resolución se encuentra disponible en:

https://github.com/harveybc/predictor/blob/master/examples/results/phase_4_1_daily/phase_4_1_cnn_1d_model_plot.png

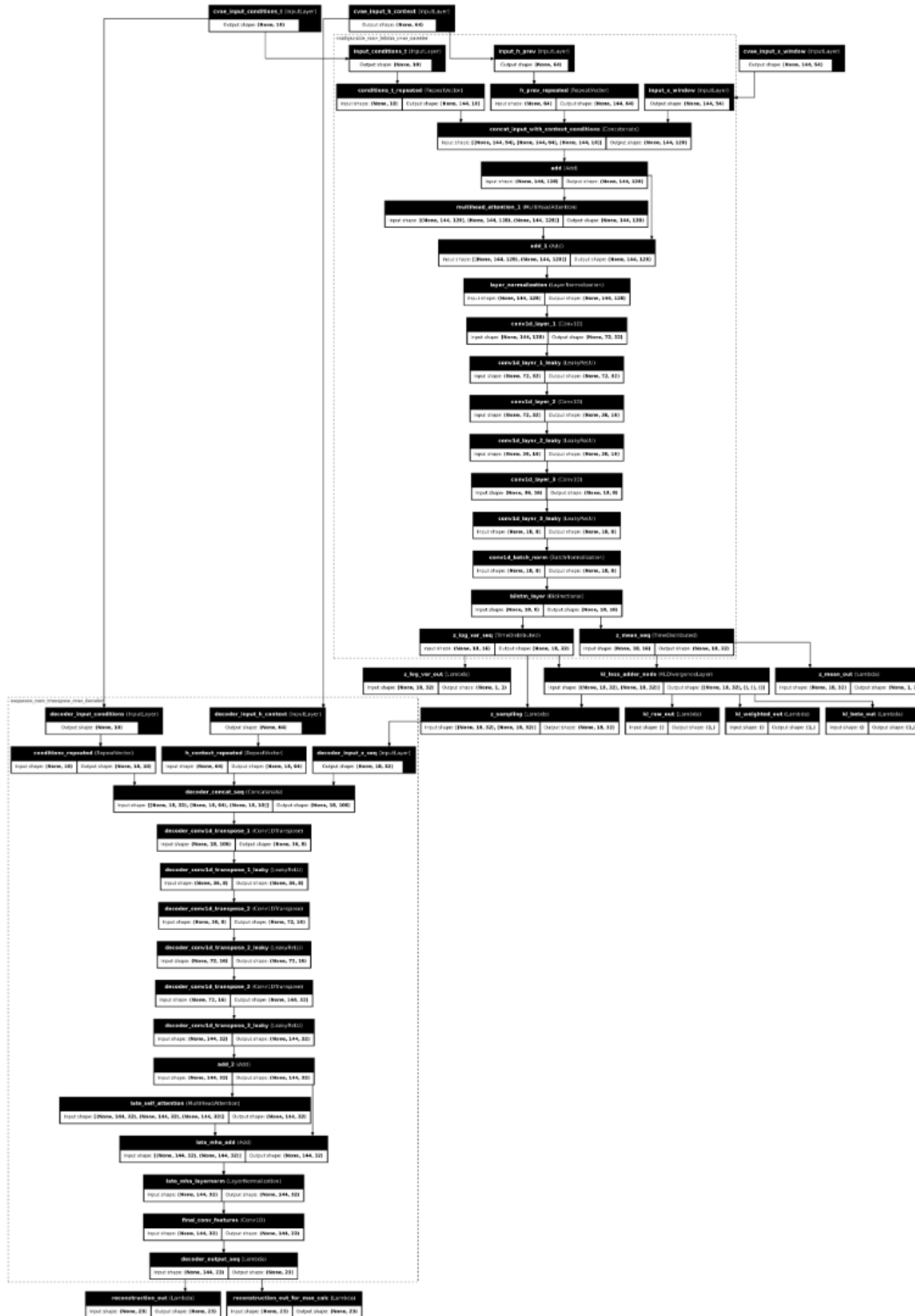
Anexo F. Cabezales predictivos autoencoder



La imagen en alta resolución se encuentra disponible en:

https://github.com/harveybc/predictor/blob/master/examples/results/phase_4_1_daily/phase_4_1_cnn_1d_model_plot.png

Anexo G. Arquitectura de autoencoder variacional condicional.



La imagen en alta resolución se encuentra disponible en:

https://github.com/harveybc/feature-extractor/blob/master/examples/results/phase_4_3/phase_4_2_cnn_1d_model_plot_ld32.png