P329	**FICHA TÉCNICA**	
	PoC ML para Forecasting	Fecha Act.: 21/05/2025

ANÁLISIS TÉCNICO

1. Información General

Título	PoC ML para Forecasting con DeepAR	**Código del Proyecto**	P329
Aplicación	Notebook – Forecasting con DeepAR en SageMaker	**Responsable del servicio**	Fernando Angarita
Elaborado por	Luis Huarcaya		
Fecha de Versión	21/05/2025	**Versión**	1.0.1

2. Control de versiones

Versión	Fecha de la versión	Descripción del cambio
1.0.1	21/05/2025	Detalle de solución técnica, observaciones y recomendaciones, optimización de hiperparámetros, validación del modelo con 15 productos.

3. Detalle de la Solución Técnica

El solución se ha desarrollado con las siguientes especificaciones técnicas:

Modelo de Forecasting DeepAR

- **Personalización de modelo** con horizonte de predicción de 6 meses y frecuencia mensual, aplicado al target 'cantidad vendida' de 15 productos seleccionados de la tienda 5503, con expansión posterior a 27 productos para demostrar escalabilidad.
- **Entrenamiento del modelo,** con 4 configuraciones, (modelo mensual con data original, modelo mensual con data original aplicada a negative binomial, modelo mensual con data modificada, modelo mensual con data modificada aplicada a negative binomial), escogiendo modelo mensual con data modificada como la configuración de mejor rendimiento (RMSE).
- **Optimización avanzada de hiperparámetros** mediante Hyperparameter Tuning Jobs en SageMaker, evaluando configuraciones de:

- Learning rate: [0.001, 0.01, 0.1]
- Epochs: [50, 100, 200]
- Num layers: [2, 3, 4]
- Dropout rate: [0.1, 0.2, 0.3]

Características Dinámicas Implementadas

- **Vector V1**: Identificador binario para período especial (2023-09-10 a 2023-11-02) que captura "gap" de mes y medio para los 15 productos
- **Características temporales**: day, weekday, week, month, quarter para capturar estacionalidad múltiple.
- **Características categóricas**: Identificadores por producto de "Tipo_Producto", "segmento_producto", "supergrupo_producto"," grupo_producto"," subgrupo_producto".

Validación y Evaluación

- **Métrica principal**: Porcentaje de pronósticos dentro del rango 90-110% del valor real, definida por el cliente como criterio de aceptación.
- **Métricas complementarias**: MAE, RMSE, MAPE para evaluación técnica comprehensiva.
- **Validación temporal**: Entrenamiento con exclusión de últimos 6 meses para evaluación en datos no vistos.
- **Cross-validation**: Validación cruzada con ventanas deslizantes para robustez del modelo Prophet.

Resultados de Rendimiento

- **DeepAR vs Prophet**: Métrica del cliente: del total de las predicciones Deepar logró clasificar 27.8% EXCELENTE, 20% BUENO,16.7% ACEPTABLE y 35.6% NECESITA MEJORA. Prophet logró 3.3% EXCELENTE, 10% BUENO, 12.2% ACEPTABLE, 74.4% NECESITA MEJORA.
- **Escalabilidad demostrada**: El entrenamiento conjunto de 27 productos con data diaria se mejoró la predicción, se redujo el RMSE de 2.9762 a 2.9011
- **Modelo diario**: Implementado para mayor granularidad, distingue períodos de no exhibición, con incremento de 2x en tiempo computacional para 15 productos.

Mejoras Implementadas

- **Preprocesamiento robusto**: Aproximación de cantidades a enteros, manejo de valores faltantes, normalización temporal.

- **Ingeniería de características**: Creación automática de vectores temporales a partir de índices de fecha.
- **Despliegue del modelo**: Desde carga de datos hasta despliegue del endpoint, completamente en AWS SageMaker.

4. Descripción de solución AWS Implementada

La solución implementada consiste en un flujo de Machine Learning en AWS SageMaker que realiza las siguientes tareas:

Componentes Principales

4.1 Notebook local (. ipynb) EDA y FE

- **Funcionalidades**:
  - Limpieza y preparación de datos con pandas y numpy
  - Ingeniería de características temporales automatizada
  - Análisis exploratorio de datos (EDA) con visualizaciones
  - Configuración de modelos DeepAR y Prophet
  - Comparación de métricas y selección de modelos

4.2 Amazon S3 Storage

- **Buckets especializados**:
  - s3://forecasting-mensual-15-v1/lilipink/data/: Datos originales - 15 productos – mensual.
  - s3://forecasting-mensual-15-v2/lilipink/data/: Datos originales – negative binomial – 15 productos – mensual.
  - s3://forecasting-mensual-15-v3/lilipink/data/:Datos modificados – 15 productos - mensual.
  - s3://forecasting-mensual-15-v4/lilipink/data/:Datos modificados – negative binomial – 15 productos - mensual.
  - s3://forecasting-mensual-27-v1/lilipink/data/:Datos modificados – 27 productos – mensual.
  - s3://forecasting-diario-27-v1/lilipink/data/:Datos modificados -27 productos – diario.
  - s3://forecasting-mensual-15-v1/lilipink/output/:Resultados – modelo 15-v1 y 15-v2 mensual.
  - s3://forecasting-mensual-15-v3/lilipink/output/:Resultados -modelo 15-v3 y 15-v4 mensual.
  - s3://forecasting-mensual-27-v1/lilipink/output/:Resultados -modelo-27-v1-mensual.
  - s3://forecasting-diario-27-v1/lilipink/output/:Resultados modelo 27-v1 diario.

4.3 SageMaker Training Jobs

- **Instancias de entrenamiento**: ml.c4.2xlarge (8 vCPU, 15 GBi)

Configuración de entrenamiento:

- Paralelización con 2 jobs simultáneos para hypertunning.
- Hyperparameter Tuning con 20 configuraciones diferentes
- Early stopping para optimización de recursos (40), prevención de overfitting.

Training Jobs

- "lilipink-forecasting-2025-05-22-15-32-57-063" modelo-27-v1-diario
- "lilipink-forecasting-2025-05-21-15-46-54-277" modelo-15-v3-mensual
- "forecasting-deepar-250521-1818-009-87cebbdc" (hyperparameter tunning job "forecasting-deepar-250521-1818") modelo-15-v4-mensual.
- "lilipink-forecasting-2025-05-22-03-48-30-015" modelo-15-mensual-final
- "forecasting-deepar-250521-1818-009-87cebbdc"
- 'lilipink-forecasting-2025-05-22-03-48-30-015' modelo-27-mensual-final

Tiempo de entrenamiento

- modelo-27-v1-diario: 22min 53s
- modelo-15-v3-mensual: 9min 15s

-Hyperparameter tunning modelo-15-v4-mensual: 2 h 38 min

- modelo-15-mensual-final:16min
- modelo-27-v1-mensual: 10min 22s RMSE:27.002529
- modelo-27-mensual-final: 16min

Parámetros Hypertuning

Name	Type	Value
mini_batch_size	Integer	449
num_cells	Integer	56
_tuning_objective_metric	FreeText	test:RMSE
context_length	FreeText	18
early_stopping_patience	FreeText	40
epochs	FreeText	400
learning_rate	FreeText	0.001
likelihood	FreeText	student-T
prediction_length	FreeText	6
time_freq	FreeText	M

P329	**FICHA TÉCNICA**	
	PoC ML para Forecasting	Fecha Act.: 21/05/2025

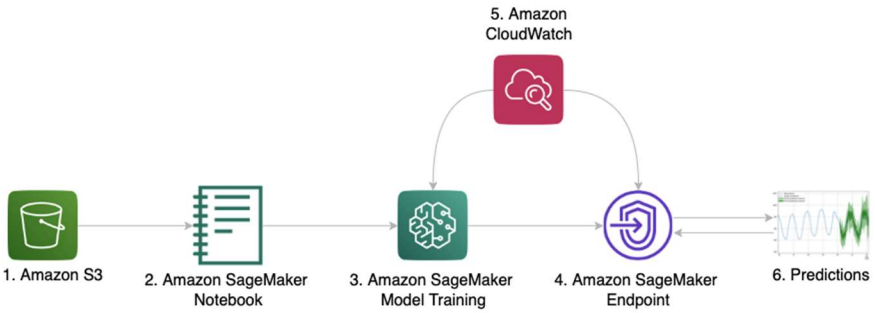
Name	Status	Final objective metric value	Creation time	Training Duration
forecasting-deepar-250521-1818-010-daabdd61	Completed	29.610124588012695	5/21/2025, 7:25:30 PM	23 minute(s)
forecasting-deepar-250521-1818-009-87cebbdc	Completed	27.00252914428711	5/21/2025, 7:24:30 PM	13 minute(s)
forecasting-deepar-250521-1818-008-fbc42c29	Stopped	36.485069274902344	5/21/2025, 7:13:11 PM	12 minute(s)
forecasting-deepar-250521-1818-007-f3a32404	Completed	34.693424224853516	5/21/2025, 7:05:38 PM	17 minute(s)
forecasting-deepar-250521-1818-006-3b8b7b54	Completed	36.48020935058594	5/21/2025, 6:59:12 PM	13 minute(s)
forecasting-deepar-250521-1818-005-21f6b659	Stopped	36.06330490112305	5/21/2025, 6:48:38 PM	10 minute(s)
forecasting-deepar-250521-1818-004-889b6143	Completed	32.92415237426758	5/21/2025, 6:45:18 PM	19 minute(s)
forecasting-deepar-250521-1818-003-89cd4a01	Completed	35.772090911865234	5/21/2025, 6:36:25 PM	8 minute(s)
forecasting-deepar-250521-1818-002-b59e07c2	Completed	32.69813537597656	5/21/2025, 6:18:53 PM	15 minute(s)
forecasting-deepar-250521-1818-001-8d581eb5	Completed	35.58380889892578	5/21/2025, 6:18:52 PM	29 minute(s)

4.4 SageMaker Endpoint

- Capacidad para alojar hasta 5 modelos simultáneamente
- Instancia: ml.m5. large (2 vCPU, 8 GB RAM)
- Tiempo de deploy
  - modelo diario 27 productos: 7min 5.5s
  - modelo mensual 27 productos: 5min 4.1s
- Tiempo de respuesta: 1.6s – 1 material
  8.6s – 27 materiales

Diagrama de Arquitectura Entrenamiento y despliegue

El siguiente diagrama ejemplifica el proyecto desarrollado, en el paso 2. se realizó un Notebook en local para el procesamiento de la data.



IAM role:
Se utilizó el siguiente rol por defecto para la implementación.
"arn:aws:iam::844598627082:role/service-role/AmazonSageMaker-ExecutionRole-20250513T105052"

5. Objetos de Aplicación

5.1 Componentes de Entrenamiento

Lista de dependencias principales

Ítem	Paquete /componente	Versión	Plataforma	Descripción	Uso
1	sagemaker	2.243.3	Python	SDK de AWS SageMaker	Entrenamiento y despliegue
2	boto3	1.38.4	Python	SDK de AWS	Interacción con servicios AWS
4	pandas	2.2.3	Python	Manipulación de datos	Procesamiento
5	numpy	1.26.4	Python	Computación numérica	Procesamiento
6	matplotlib	3.10.1	Python	Visualización	Gráfico de series temporales
7	seaborn	0.13.2	Python	Visualización	Análisis exploratorio
8	openpyxl	3.1.5	Python	Excel	Excel
9	tqdm	4.67.1	Python	Monitoreo	Monitoreo

Notebook desarrollado con Python 3.12

5.2 Variables de configuración

Variables de entrenamiento

Ítem	Nombre	Valor Hypertuning	Descripción
1	PREDICTION_LENGTH	6	Horizonte de predicción en meses
2	CONTEXT_LENGTH	18	Longitud de contexto para el modelo
4	FREQ	'M'	Frecuencia de las series temporales
5	EPOCHS	400	Número de épocas de entrenamiento
6	LEARNING_RATE	0.001	Tasa de aprendizaje
7	BATCH_SIZE	449	Tamaño del lote
8	NUM_LAYERS	2	Número de capas LSTM
9	NUM_CELLS	56	Número de células
10	LIKELIHOOD	STUDENT-T	Modelo probabilístico

Variables de endpoint

Ítem	Nombre	Valor por defecto	Descripción
1	ENDPOINT_INSTANCE_TYPE	'ml.m5. large'	Tipo de instancia para endpoint
2	INITIAL_INSTANCE_COUNT	1	Número inicial de instancias
3	MAX_CONCURRENT_TRANSFORMS	10	Transformaciones concurrentes máximas
4	MODEL_SERVER_TIMEOUT	60	Timeout del servidor en segundos

8. Métricas de Rendimiento y Escalabilidad

8.1 Métricas del Modelo

Métrica Cliente

Pronóstico Prophet

A continuación, se muestra algunos probatorios de las predicciones, el archivo completo será compartido en el sharepoint del cliente.

Material	Fecha	Predicción	Real
20000337001	1/11/2024	13.3700539	22
20000337001	1/12/2024	41.6506319	41
20000337001	1/01/2025	20.0863052	17
20000337001	1/02/2025	10.8655244	10
20000337001	1/03/2025	19.5959672	14
20000337001	1/04/2025	14.8386539	3

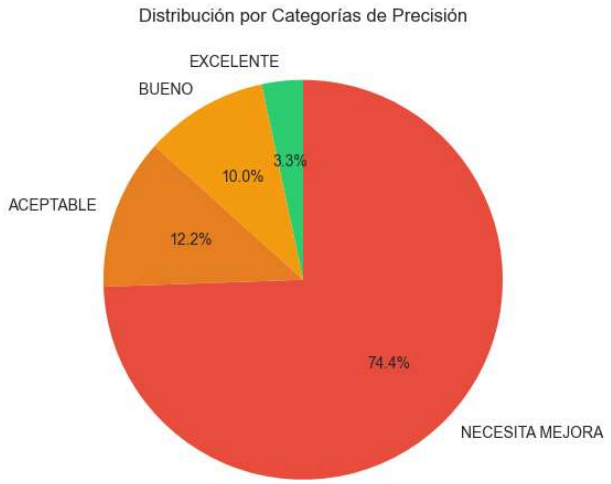
P329	**FICHA TÉCNICA**	
	PoC ML para Forecasting	Fecha Act.: 21/05/2025

20000400003	1/11/2024	10.1143487	4
20000400003	1/12/2024	21.1202243	14
20000400003	1/01/2025	12.5384443	5
20000400003	1/02/2025	10.2650815	4
20000400003	1/03/2025	6.74359975	1
20000400003	1/04/2025	7.44139191	1

20000815002	1/11/2024	45.1977426	25
20000815002	1/12/2024	38.0156204	34
20000815002	1/01/2025	14.5159021	6
20000815002	1/02/2025	8.25130099	11
20000815002	1/03/2025	27.5406353	4
20000815002	1/04/2025	15.2724635	13

Criterio

```
if 95 <= porcentaje <= 105:
        return 'EXCELENTE'
    elif 90 <= porcentaje <= 110:
        return 'BUENO'
    elif 80 <= porcentaje <= 120:
        return 'ACEPTABLE'
    else:
```



P329	**FICHA TÉCNICA**	
	PoC ML para Forecasting	Fecha Act.: 21/05/2025

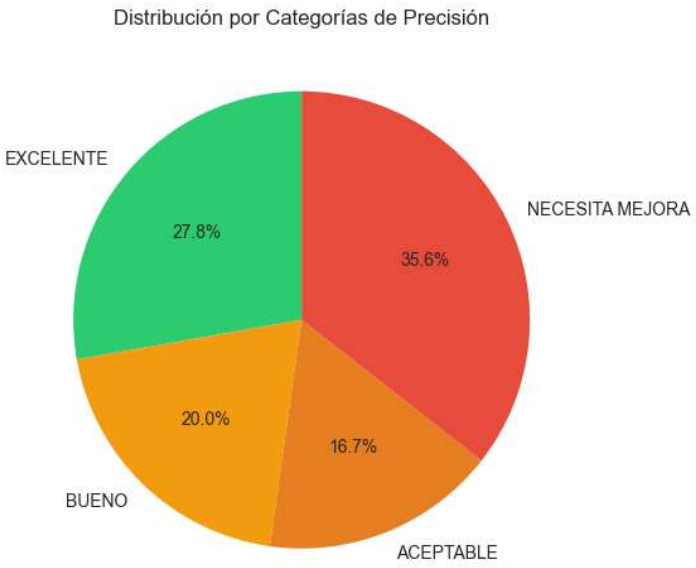
Métrica Cliente
Pronóstico DeepAR

Fecha	Material	0.1	0.5	0.9	Real
2024-11-01	20000337001	29.75198936	33.1474152	36.35353	33
2024-12-01	20000337001	40.60406876	44.7514839	48.56557	44
2025-01-01	20000337001	16.1166935	17.3929882	19.1844	17
2025-02-01	20000337001	9.22453022	10.6521606	12.38436	10
2025-03-01	20000337001	6.3241539	11.0567541	16.00208	14
2025-04-01	20000337001	2.13414669	7.17127132	11.0784	4
2024-11-01	20000400003	3.511388302	5.63880157	7.35116	5
2024-12-01	20000400003	10.94414139	13.9024096	17.06828	14
2025-01-01	20000400003	4.117705345	4.82056952	5.672948	5
2025-02-01	20000400003	2.737850905	3.51785421	4.493963	4
2025-03-01	20000400003	2.222531319	4.48396111	6.030576	1
2025-04-01	20000400003	-1.04762828	1.31956029	3.686026	2

Criterio

```
if 95 <= porcentaje <= 105:
        return 'EXCELENTE'
    elif 90 <= porcentaje <= 110:
        return 'BUENO'
    elif 80 <= porcentaje <= 120:
        return 'ACEPTABLE'
    else:
```

P329	**FICHA TÉCNICA**	
	PoC ML para Forecasting	Fecha Act.: 21/05/2025



Métrica RMSE, MAE, MAPE

Se calcularon las métricas para un modelo mensual de 27 productos sin procesar.

Prophet

'metricas': {'RMSE': 176.89108018624384,
'MAE': 59.59196349600039,
'MAPE': 1672.7402603632004},

DeepAR

'metricas': {'RMSE': 51.43030962591935,
'MAE': 26.346820054412962,
'MAPE': 179.72815758201594},

Se observa que DeepAR tiene mejores métricas de testing.
Posteriormente se modificó la data y se realizo hypertuning en el modelo mensual de 27 productos lograndose disminuir el RMSE a 27.002529.

9. Próximos Pasos y Recomendaciones

9.1 Mejoras Técnicas

Incorporar características externas:
Promociones, eventos especiales, días festivos
Detección de periodos de no exhibición:
Identificación de estos periodos y días con venta de 0 unidades.
Ensemble models:
Combinación de DeepAR con otros algoritmos para mayor robustez
Feature importance:
Análisis de importancia de características temporales

9.2 Escalabilidad

Modelo multi-tienda:
Extensión a múltiples tiendas con características específicas
Pipeline automatizado:
CI/CD para reentrenamiento automático
Real-time inference:
Capacidad de predicción en tiempo real
Edge deployment:
Despliegue en edge para latencia ultra-baja

9.3 Monitoreo y Mantenimiento

Model drift detection:
Detección automática de degradación del modelo
A/B testing framework:
Comparación continua de versiones de modelo
Automated retraining:
Re-entrenamiento automático basado en nuevos datos
Performance dashboards:
Dashboards en tiempo real para monitoreo

5. Aprobado por:

Nombres y Apellidos	***Cargo / Función***	***Firma***
Luis Huarcaya	ML Engineer	