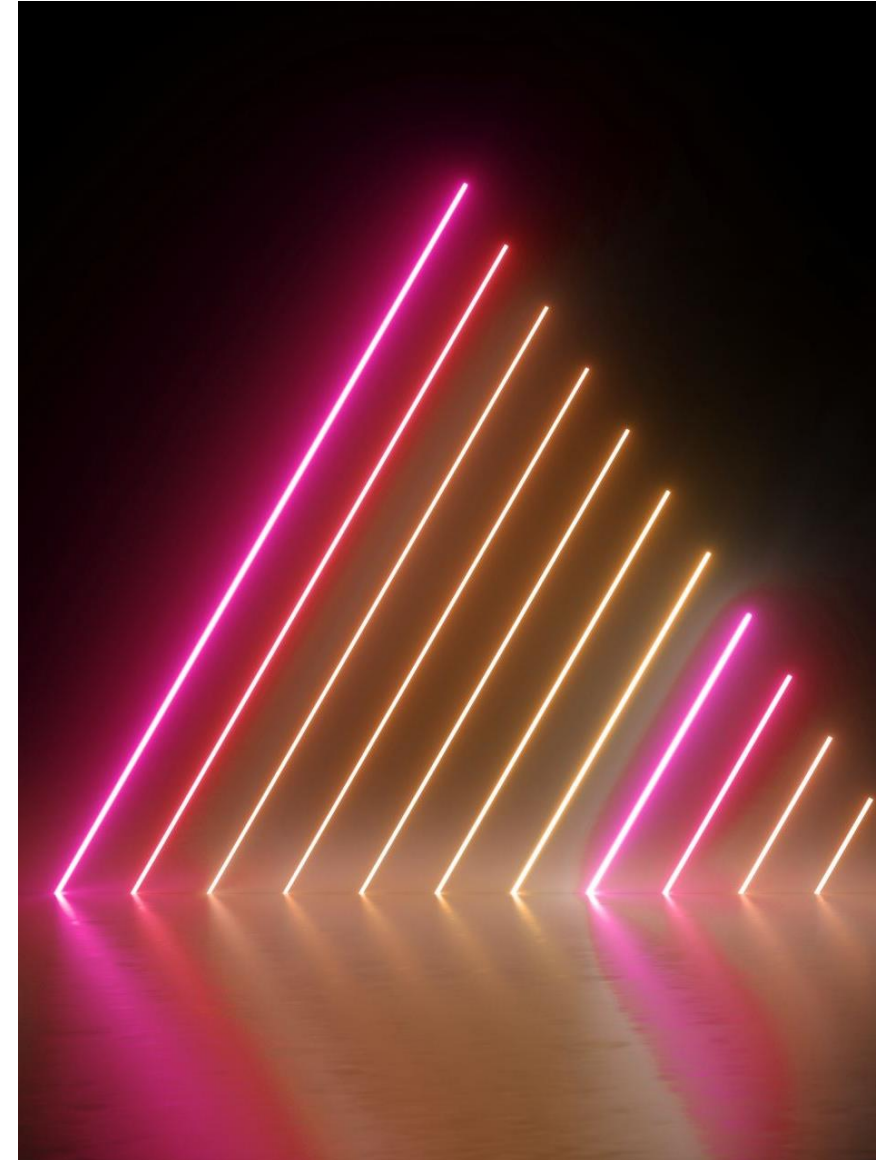


ANÁLISIS COMPARATIVO DE MODELOS PREDICTIVOS  
DE MACHINE LEARNING PARA LA PROYECCIÓN DE  
VENTAS DE PRODUCTOS NESTLÉ EN  
SUPERMERCADOS TIA EN ECUADOR

**MAESTRÍA EN INTELIGENCIA DE NEGOCIOS Y CIENCIA  
DE DATOS**

**Flores V, Guerra E.**

**Agosto 24,2023**



# Contenido

- Introducción
- Marco Teórico
- Revisión de la Literatura
- Análisis descriptivo de variables
- Descripción de los modelos
- Resultados de los modelos
- Conclusiones
- Recomendaciones
- Referencias





Nestlé

# Introducción

- Nestlé Ecuador es una empresa líder en la industria de bienes de consumo que maneja un gran volumen de datos de ventas generados diariamente a través de sus distintos canales. Uno de sus principales clientes es la cadena de supermercados TIA. Actualmente, Nestlé Ecuador utiliza los datos de ventas de TIA para informar las estrategias de ventas y marketing, así como para optimizar la cadena de suministro.



## Pregunta central de investigación

- ¿Cómo pueden Nestlé Ecuador y su cliente TIA beneficiarse de la implementación de modelos predictivos de machine learning para proyectar las ventas futuras de manera efectiva, y cuál de los dos modelos propuestos, Random Forest o Redes Neuronales, es más adecuado para esta tarea?





# Objetivo

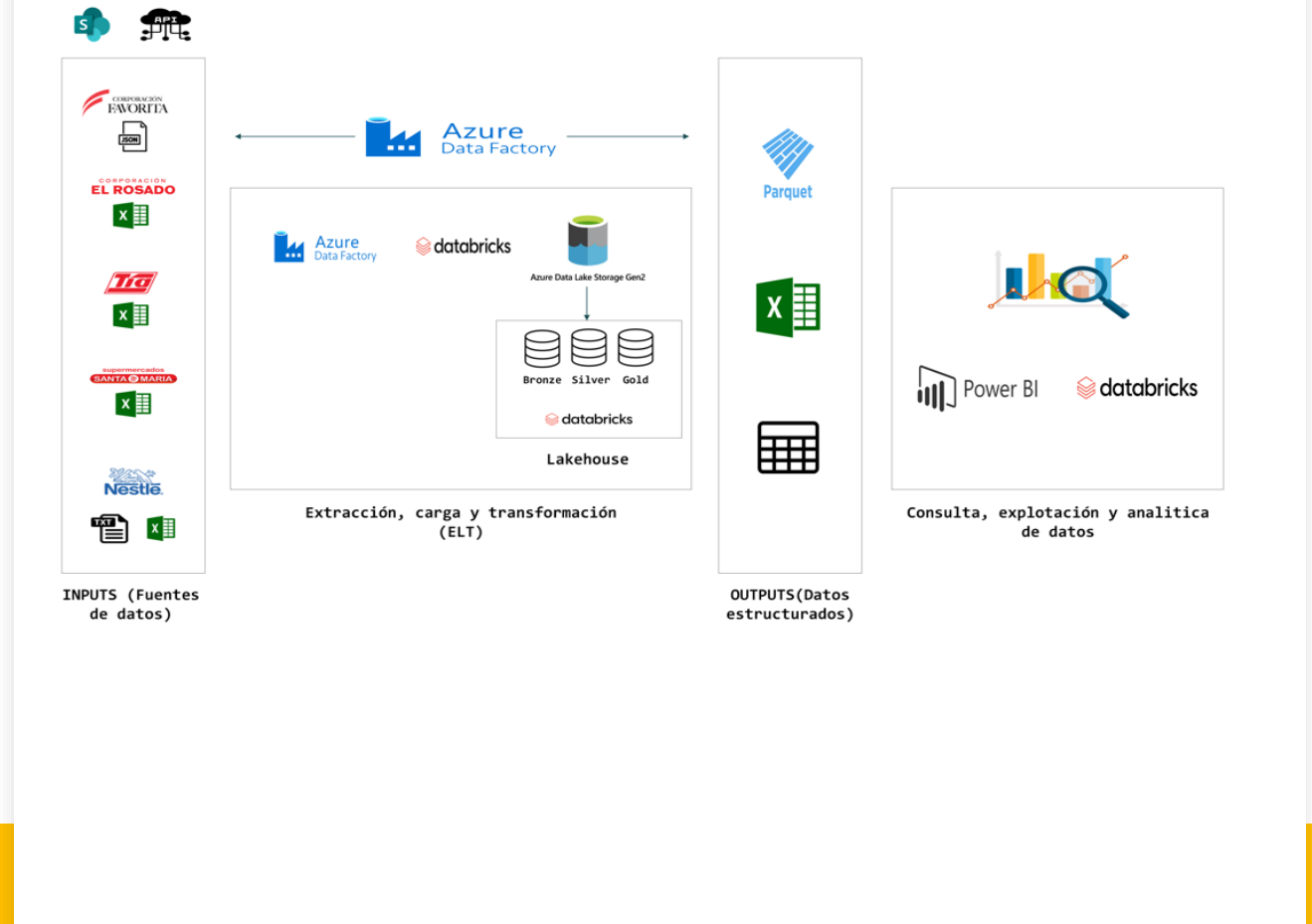
---

- Desarrollar y aplicar modelos de Redes Neuronales y Random Forest para la proyección de ventas de los productos Nestlé en los supermercados TIA en Ecuador, con el propósito de comparar la eficacia de estos dos métodos de Machine Learning en la precisión de las predicciones de ventas y en la identificación de los factores más influyentes en dichas ventas, proporcionando así una herramienta efectiva para la toma de decisiones y la planificación estratégica en la industria de bienes de consumo.



# Recolección de Datos

- Arquitectura de datos de Nestle Ecuador enfocada en Big Data para ventas y stock de clients de todos los canales.



# EDA (Exploratory Data Analysis)

- Índices de correlación.
- Eliminación de variables con alto índice de correlación.
- Definición de variable dependiente: totalSo.
- Numero de registros: 188490
- Rango de fechas: Enero a Julio 2023

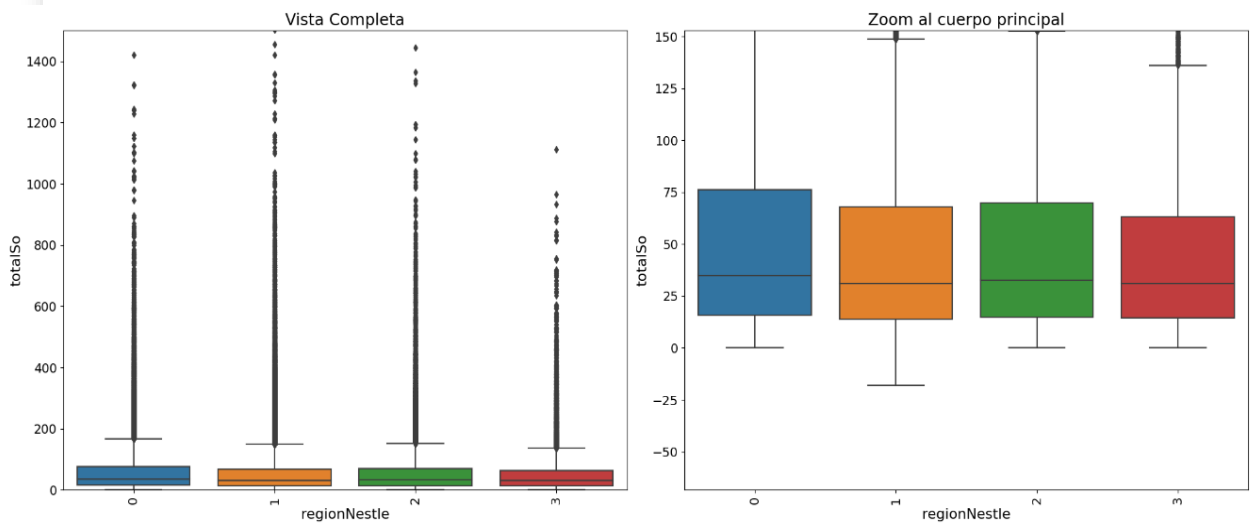
	mes	unidadesVendidas	ventaNeta	totalSo	stock	totalStock
mes	1.000000	0.052599	0.094477	0.099716	0.008117	0.022849
unidadesVendidas	0.052599	1.000000	0.761121	0.765171	0.577638	0.262593
ventaNeta	0.094477	0.761121	1.000000	0.984587	0.488755	0.471933
totalSo	0.099716	0.765171	0.984587	1.000000	0.455794	0.445920
stock	0.008117	0.577638	0.488755	0.455794	1.000000	0.696380
totalStock	0.022849	0.262593	0.471933	0.445920	0.696380	1.000000

nombre	tipo de dato	tipo de variable	descripción
anio	date	categorica	Año de la venta
mes	int	categorica	Mes de la venta
localNestle	string	categorica	Nombre de local de TIA donde se hizo la venta
ciudadNestle	string	categorica	Nombre de la ciudad donde se encuentra el local
regionNestle	string	categorica	Nombre de la región donde se encuentra la ciudad
subcadenaNetle	string	categorica	Nombre de la subcadena a la que corresponde el local
unidadesVendidas	double	numérica	Cantidad vendida en unidades
ventaNeta	double	numérica	Venta neta reportada por TIA
totalSo	double	numérica	Venta realizada por Nestlé a TIA
stock	double	numérica	Cantidad en unidades de stock
totalStock	double	numérica	Stock valorizado según precios de venta de Nestlé a TIA

**Nota:** No se encuentran datos faltantes debido al proceso previo de transformación.

# EDA (Exploratory Data Analysis)

- Identificación de outliers o datos atípicos.
- Transformación de variables categóricas a numéricas mediante label encoder



regionNestle	mean	25%	50%	75%	std
COSTA	65,014094	15,83199	34,92	76,103412	91,77073
GUAYAS	59,219247	13,8166	31,106495	67,7739	95,375599
SIERRA NORTE	60,888669	14,76072	32,628536	69,869666	87,318081
SIERRA SUR	53,800275	14,4	30,957696	63,0504	73,188151

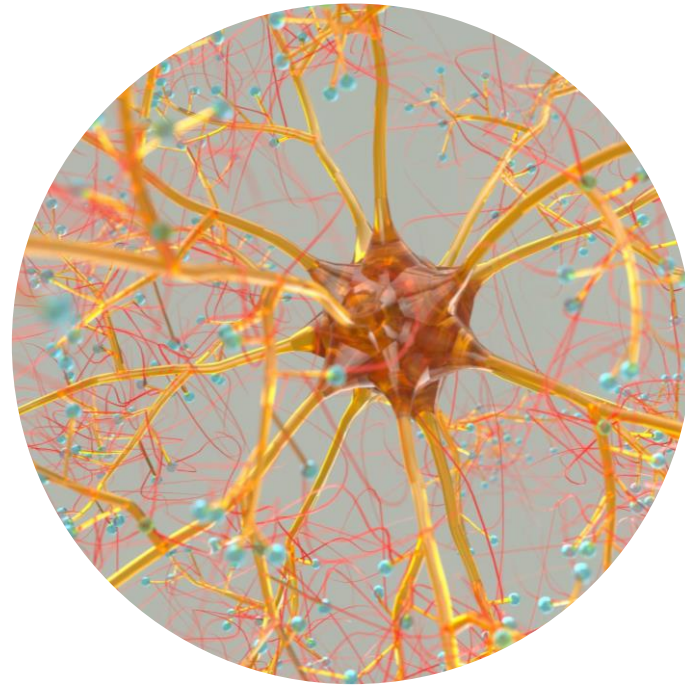
**Nota:** Se evidencian valores de sell out negativos.



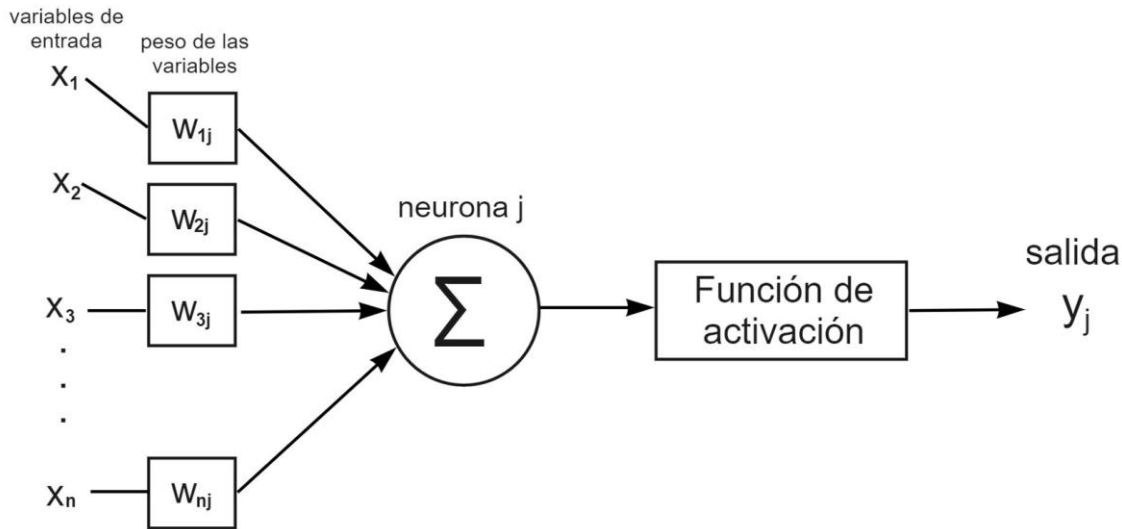
# Redes Neuronales Artificiales

## Ideas Intuitivas

- Las redes neuronales artificiales (ANN), son modelos **estadísticos computacionales** que se inspiran en el **cerebro humano**.
- Están formadas por diferentes **nodos** que funcionan como neuronas, y **que transmiten señales** e información entre sí.
- Estas redes reciben **diferente información de entrada**, la procesan en conjunto y **generan una salida con las predicciones establecidas en función** de lo que se haya programado.



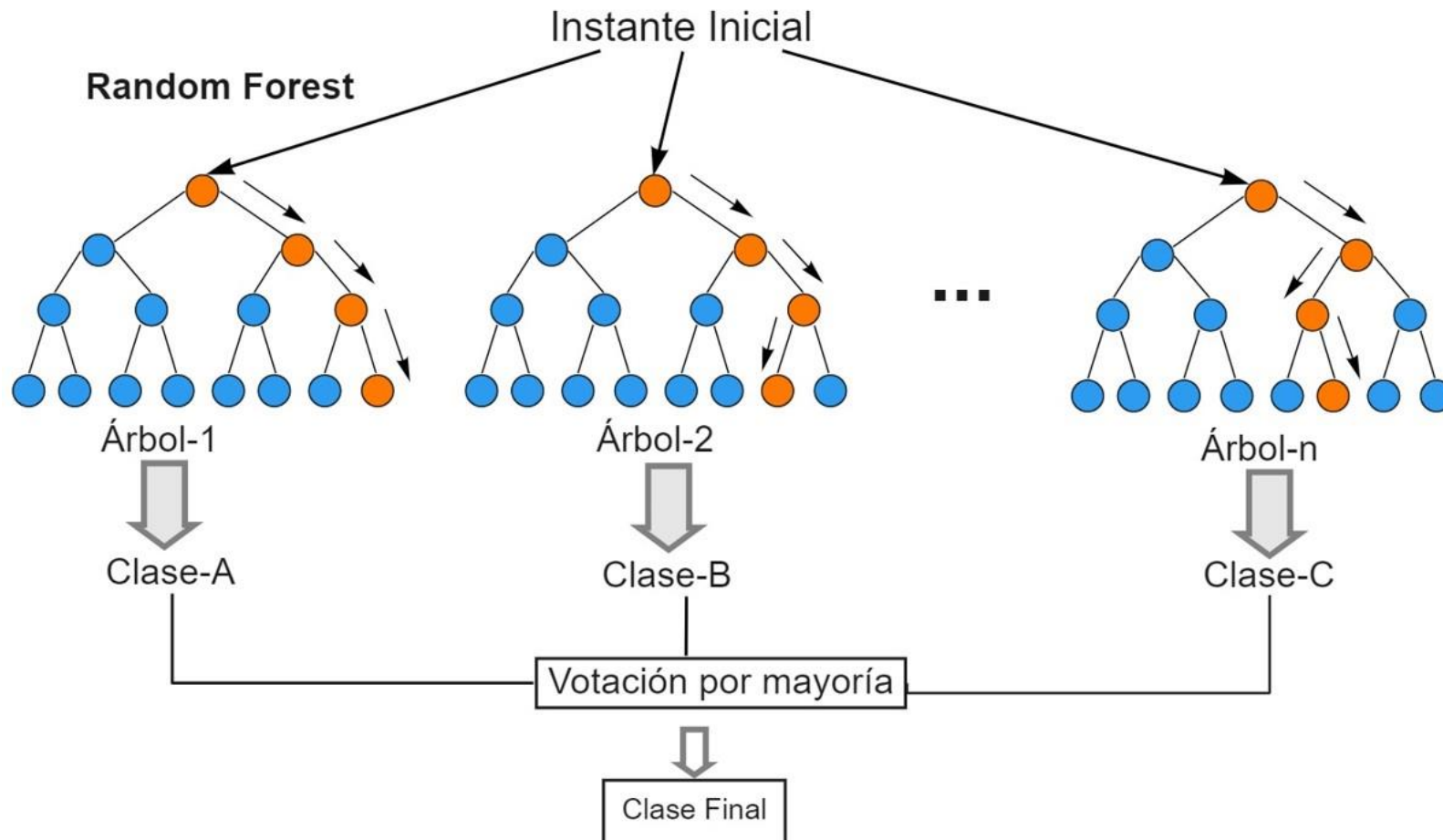
# Redes Neuronales Artificiales



Fuente: Recuperado de WikiStats (2020)

- La entrada para cada neurona en la primera capa oculta de la red es la suma ponderada de todas las conexiones entre la capa de entrada y la neurona en la capa oculta. Esta suma ponderada es llamada a veces estímulo o entrada de red.
- El estímulo de red se transforma por la activación de la neurona o de la función de transferencia  $f(E)$ , para producir un nuevo valor de salida de la neurona.
- Además de los estímulos de red, el bias ( $\mu$ ) se añade generalmente para compensar la entrada

## Random Forest Simplificado



Fuente: Recuperado de Datasource (2020)

## Random Forest

Los modelos de Bosques Aleatorios están formados por un conjunto de árboles de decisión individuales, cada uno entrenado con una muestra ligeramente distinta de los datos de entrenamiento.

Coeficiente de determinación	Profundidad del árbol	Número de estimadores
97.28%	None	20
97.22%	20	20
95.45%	10	20
66.70%	3	20

Importancia Media	Importancia Estándar	Variable
109.760399	0.247757	unidadesVendidas
85.127356	0.216876	stock
75.162473	0.262026	totalStock
13.796769	0.335464	mes
6.609943	0.326726	localNestle
5.002447	0.398214	ciudadNestle
2.859698	0.32066	regionNestle

Construcción  
Random  
Forest

dense_3_input	input:	[(None, 7)]
InputLayer	output:	[(None, 7)]



dense_3	input:	(None, 7)
Dense	output:	(None, 15)



dense_4	input:	(None, 15)
Dense	output:	(None, 10)



totalSO	input:	(None, 10)
Dense	output:	(None, 1)

Construcción  
Redes Neuroanles  
Artificiales



# Resultados

Modelos	Error Cuadrático Medio (MSE)	$R^2$
Random Forest	393.51	95%
Redes Neuronales Artificiales	1278.84	84%

Modelos	Tiempo de ejecución (segundos)	Calibrado hiperparámetros (segundos)
Random Forest	1.31	450
Redes Neuronales Artificiales	95	13242



# Propuesta de solución

## 01 - Estrategia Organizacional Alineada

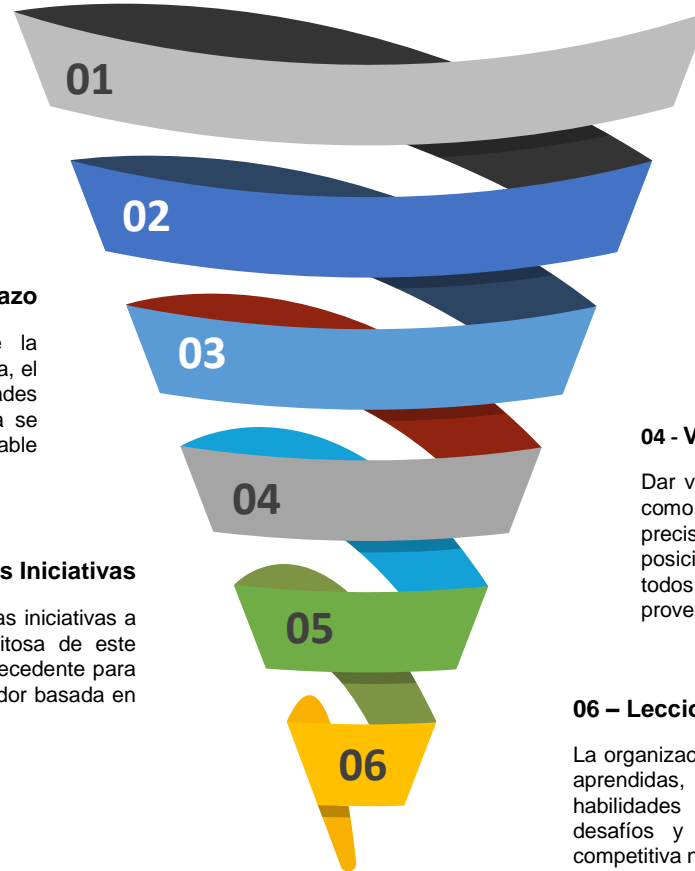
Alinear la estrategia organizacional redefiniendo la manera en que Nestlé Ecuador percibe, reacciona y se adelanta a los patrones del mercado a través del uso de los datos.

## 03 - Competitividad a Largo Plazo

Generar competitividad a largo plazo mediante la estrategia propuesta mejorando la eficiencia operativa, el entendimiento y la satisfacción de las necesidades cambiantes de sus consumidores. De esta manera se refuerza la imagen de Nestlé como una marca confiable y en sintonía con sus clientes.

## 05 - Base Sólida para Futuras Iniciativas

Tener una base sólida para futuras iniciativas a través de la implementación exitosa de este proyecto ya que se sentará un precedente para futuras iniciativas en Nestlé Ecuador basada en analítica de datos.



## 02 - Fomento de la Cultura de Innovación

Fomentar la cultura de innovación en cuanto el uso de los datos mediante la implementación y adaptación a técnicas avanzadas de análisis de datos. Una organización que es capaz de innovar y adaptarse rápidamente a cambios del mercado no solo es más rentable, sino también más confiable y atractiva para inversiones futuras.

## 04 - Valor Agregado para los Stakeholders

Dar valor agregado a los Stakeholders lo cual viene como consecuencia de mejorar la eficiencia, la precisión y la adaptabilidad. Nestlé Ecuador estará en posición de ofrecer un valor agregado significativo a todos sus stakeholders, desde clientes hasta proveedores y accionistas.

## 06 – Lecciones aprendidas

La organización podrá aprovechar las lecciones aprendidas, los datos acumulados y las habilidades desarrolladas para abordar otros desafíos y oportunidades con una ventaja competitiva notable.

# Conclusiones

## Impacto en la Toma de Decisiones Gerenciales

Con la implementación exitosa de este proyecto, la gerencia de Nestlé Ecuador está ahora más empoderada que nunca para tomar decisiones basadas en datos concretos y análisis profundos. Esta capacidad no solo optimiza las operaciones actuales, sino que también proporciona una hoja de ruta

## Innovación como Piedra Angular

Este proyecto ha reafirmado que la innovación no es un lujo, sino una necesidad para mantener la relevancia y el liderazgo en un mercado en constante cambio. La adaptación de Nestlé Ecuador a nuevas técnicas y enfoques demuestra su compromiso con la excelencia y su visión de futuro, estableciendo un precedente para futuras iniciativas innovadoras.

## Eficiencia de los modelos

Según los resultados obtenidos, el modelo Random Forest supera significativamente en precisión a las Redes Neuronales Artificiales, con un Error Cuadrático Medio (MSE) de 393.51 frente a 1278.84, y un coeficiente de determinación  $R^2$  de 95% frente a 84%. Esto indica que, para este conjunto de datos y la problemática específica de Nestlé Ecuador, el Random Forest es más adecuado para predecir las ventas con precisión.



# Nestlé

## Tiempo de ejecución

Es importante considerar la eficiencia en tiempo, especialmente en escenarios donde se requiere una respuesta rápida. En este caso, Random Forest demostró ser mucho más eficiente, requiriendo solamente 1.31 segundos para su ejecución frente a los 95 segundos requeridos por la Red Neuronal. Además, el calibrado de hiperparámetros también fue significativamente más rápido en el Random Forest.

# Recomendaciones

---

**Adoptar el modelo Random Forest:** Dadas sus ventajas en precisión, eficiencia en tiempo y robustez, Nestlé Ecuador debería considerar implementar el modelo Random Forest como herramienta principal para la predicción de sus ventas como punto de partida en la analítica de datos.

---

**Monitoreo constante:** Aunque el Random Forest ha demostrado ser superior en este análisis, es esencial monitorear constantemente su desempeño y estar abierto a evaluar y adoptar nuevas técnicas o modelos en el futuro, ya que las condiciones del mercado, los datos y las tecnologías cambian con el tiempo.

---

**Explorar nuevas variables:** Aunque las variables identificadas como 'unidadesVendidas', 'stock', 'totalStock', y 'mes' son cruciales, podría ser valioso explorar la incorporación de nuevas variables o indicadores que puedan influir en las ventas, para mejorar aún más la precisión del modelo.

---

**Continuar con la Inversión en Análisis de Datos y Herramientas de Big Data:** La adopción temprana de técnicas avanzadas de análisis de datos ha demostrado ser fructífera. Recomendamos que Nestlé Ecuador siga invirtiendo en estas herramientas y en la formación de su personal para maximizar los beneficios.

---

**Fomentar una Cultura de Innovación:** La innovación debe ser una constante en la organización. Es esencial organizar talleres, seminarios y programas de formación que mantengan al equipo actualizado y motivado.

---

# Referencias

- Dhar, V. (2012). Data Science and Prediction. *Communications of the ACM*, 56, 64-73.
- Hyndman, R., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. OTexts.
- Kelleher, J., Mac Namee, B., & D'Arcy, A. (2020). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. MIT Press.
- Kuyver, T., Ragan-Kelley, B., Brian, G., Perez, F., Granger, B., Bussonier, M., . . . Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, 87-90.
- Usme Valencia, M., & Rojas Díaz, J. (2022). *bibliotecadigital.udea.edu.co*. Obtenido de bibliotecadigital.udea.edu.co: [https://bibliotecadigital.udea.edu.co/bitstream/10495/29133/1/UsmeMateo\\_2022\\_ModelosPronosticoVentas.pdf](https://bibliotecadigital.udea.edu.co/bitstream/10495/29133/1/UsmeMateo_2022_ModelosPronosticoVentas.pdf)
- Yang, C., Lewis, G. A., Brower-Sinning, R. A., & Kästner, C. (2022). Data Leakage in Notebooks: Static Detection and Better. *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*, 1-12.
- WikiStat. (2020). *Neural networks and introduction to deep learning*.