

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ**

**FACULTAD DE CIENCIAS E INGENIERÍA**



**PUCP**

**MARCO TEÓRICO PARA TESIS DE MEJORA DEL ERROR DE PRONÓSTICO  
UTILIZANDO MÉTODOS DE MACHINE LEARNING ORIENTADO AL PROCESO  
DE VENTAS DE UNA EMPRESA DE COSMÉTICOS PERUANA**

**Trabajo de investigación para la obtención del grado de BACHILLER EN  
CIENCIAS CON MENCIÓN EN INGENIERÍA INDUSTRIAL**

**AUTOR**

Sebastian Linares Rosas

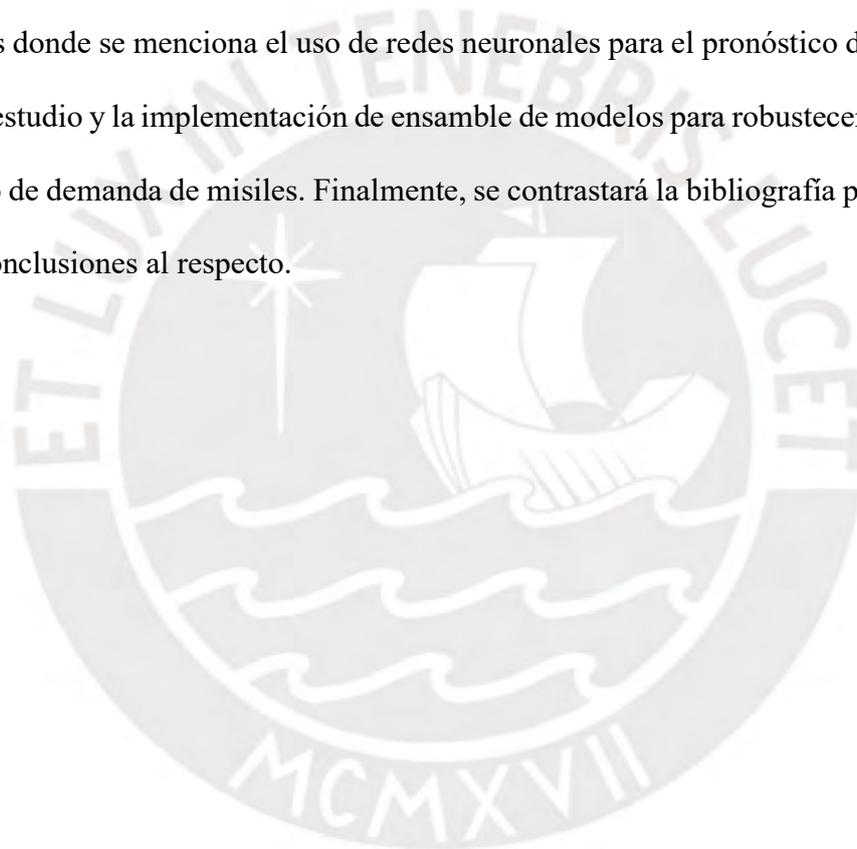
**ASESOR:**

Mery Roxana León Perfecto

Lima, setiembre, 2020

## Resumen

El presente trabajo de investigación tiene como objetivo analizar y contrastar las actuales soluciones de analítica avanzada en problemas de Supply Chain. Para esto, se empezará armando un marco teórico que consolide los conceptos fundamentales de pronósticos, machine learning y diagnóstico de procesos. Con esta base, se presentarán los puntos principales de diversas publicaciones y trabajos académicos relacionados al uso de inteligencia artificial en Supply Chain, incluyendo métodos y resultados obtenidos. Entre estos, destacan las publicaciones donde se menciona el uso de redes neuronales para el pronóstico de demanda en un centro de estudio y la implementación de ensamble de modelos para robustecer la predicción de un modelo de demanda de misiles. Finalmente, se contrastará la bibliografía presentada y se elaborarán conclusiones al respecto.



## Tabla de Contenido

Resumen .....	i
Índice de Tablas .....	iii
Índice de Figuras .....	iii
Marco Teórico.....	1
Pronósticos.....	1
Fundamentos de Machine Learning .....	4
Diagnóstico de Procesos.....	7
Contenido de la investigación .....	10
1. A systematic review of the research trends of machine learning in supply chain management .	10
2. A neural network solution for forecasting labor demand of drop-in peer tutoring centers with long planning horizons .....	11
3. Prediction of air travel demand using a hybrid artificial neural network (ANN) with Bat and Firefly algorithms: a case study .....	14
4. Urban cold-chain logistics demand predicting model based on improved neural network model .....	15
5. Supply–Demand Prediction for Agile Manufacturing with Deep Neural Network.....	16
6. Supply Chain Demand Forecasting; A Comparison of Machine Learning Techniques and Traditional Methods.....	18
7. An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain .....	19
8. Forecasting of sales by using fusion of Machine Learning techniques .....	21
9. Analysis on Machine Learning Algorithms and Neural Networks for Demand Forecasting of Anti-Aircraft Missile Spare Parts .....	22
10. Analysis of book sales prediction at Amazon marketplace in India: a machine learning approach .....	23
Discusión.....	25
Conclusiones.....	28
Bibliografía.....	30

## Índice de Tablas

Tabla 1 Resumen de resultados al comparar ANN y regresión .....	14
Tabla 2 Resumen de resultados al comparar SVM, ANN y diversos modelos clásicos .....	18
Tabla 3 Resumen de resultados al comparar combinaciones de ARIMA, ARNN, XGBoost y SVM.....	22
Tabla 4 Resumen de resultados al comparar XGBoost, MLP y diversos modelos de machine learning .....	23

## Índice de Figuras

Figura 1 Ejemplos de tendencia de pronosticos.....	3
Figura 2 Estructura de un perceptrón .....	5
Figura 3 Estructura de una red neuronal simple .....	5
Figura 4 Gráfica de la función ReLU .....	6
Figura 5 Ejemplo de Espina de Ishikawa .....	7
Figura 6 Diferencias entre distintos niveles de correlación.....	8
Figura 7 Distribución de los algoritmos encontrados en las investigaciones .....	11
Figura 8 Funcionamiento del algoritmo NNSOA .....	13
Figura 9 Gráfico de líneas de valores reales y predicciones con BP y PCA-BP .....	15
Figura 10 Diferencias de tiempo operacional entre BP y PCA-BP.....	16
Figura 11 Diagrama de flujo de la arquitectura de una red neuronal que utiliza Feature Engineering	17
Figura 12 Diagrama de flujo de una arquitectura que combina Deep Learning, SVM y series de tiempo .....	20

## **Marco Teórico**

El acelerado desarrollo de múltiples disciplinas tecnológicas como el internet de las cosas, la computación en la nube y la nanotecnología es la característica principal de la “4ta revolución industrial” (Skilton y Hovsepian, 2018, p. 9-15). Como lo describe el fundador del Foro Económico Mundial, esta transformación se está llevando a cabo de forma disruptiva en todo el mundo, y ha permitido el impulso de soluciones innovadoras que benefician tanto a consumidores como al sector empresarial (Schwab, 2016). Uno de los campos de investigación que ha sido favorecido con lo anterior mencionado es la inteligencia artificial. Este tema ha sido tratado desde hace más de tres décadas, con John Hopfield como uno de los pioneros, quien en 1982 publicó un artículo sobre las propiedades de las redes neuronales artificiales y su capacidad para generalizar patrones (Hopfield, 1982). Sin embargo, la creciente oferta de información disponible producto de la era de la digitalización, en conjunto con el incremento de recursos computacionales, son los factores que han permitido que esta tecnología se vuelva relevante y brinde herramientas efectivas para cumplir con tareas que involucren predicción en base a datos (Gupta, et al., 2017, p. 1). En la presente investigación se analizarán los casos de aplicación de algoritmos de inteligencia artificial en un ámbito en específico: la planificación de demanda y Supply Chain. Asimismo, se abordarán métodos de diagnóstico que ayuden a determinar la necesidad de aplicar una solución orientada a mejorar pronósticos y se comentarán los fundamentos teóricos que respaldan el funcionamiento de los modelos predictivos

### **Pronósticos**

En el libro de Administración de Operaciones de Krajewski, se define un pronóstico como una predicción de acontecimientos futuros que se utiliza con propósitos de planificación

(Krajewski, Malhorta y Ritzman, 2018, p. 277). Asimismo, se menciona la importancia de estos al ser el punto de partida de procesos de planificación de servicios, producción, inventarios e ingresos. Por lo tanto, un incorrecto método de pronóstico podría perjudicar las operaciones de diversas áreas de la empresa, por lo que, de ser así, deberá identificarse el problema a tiempo con el objetivo de reducir las pérdidas relacionadas a este.

Por otro lado, Chase (2018) comenta que existen diferentes enfoques a tomar en cuenta dependiendo del tipo de la finalidad del pronóstico que se requiere utilizar. Este autor define dos categorías de pronósticos: estratégicos y tácticos. Los primeros son de ayuda para determinar la demanda agregada a lo largo del año y para realizar estimaciones a largo plazo, lo cual se requiere en la determinación de la estrategia, el cálculo de la capacidad y la planeación de operaciones y ventas, por ejemplo. La segunda categoría se refiere a pronósticos que se usan en actividades del día a día, como la gestión de inventarios, la compra de suministros y la planeación de producción; para todas estas se necesitan valores a corto plazo (2018, p. 445). Por lo tanto, se debe establecer claramente los objetivos del pronóstico que se quiere elaborar para obtener resultados satisfactorios.

Asimismo, dependiendo del tipo de productos que maneje la empresa a analizar, se observarán distintos patrones de demanda en los resultados históricos. Krajewski et al. (2018) incluye los siguientes gráficos para ejemplificar lo anteriormente mencionado:

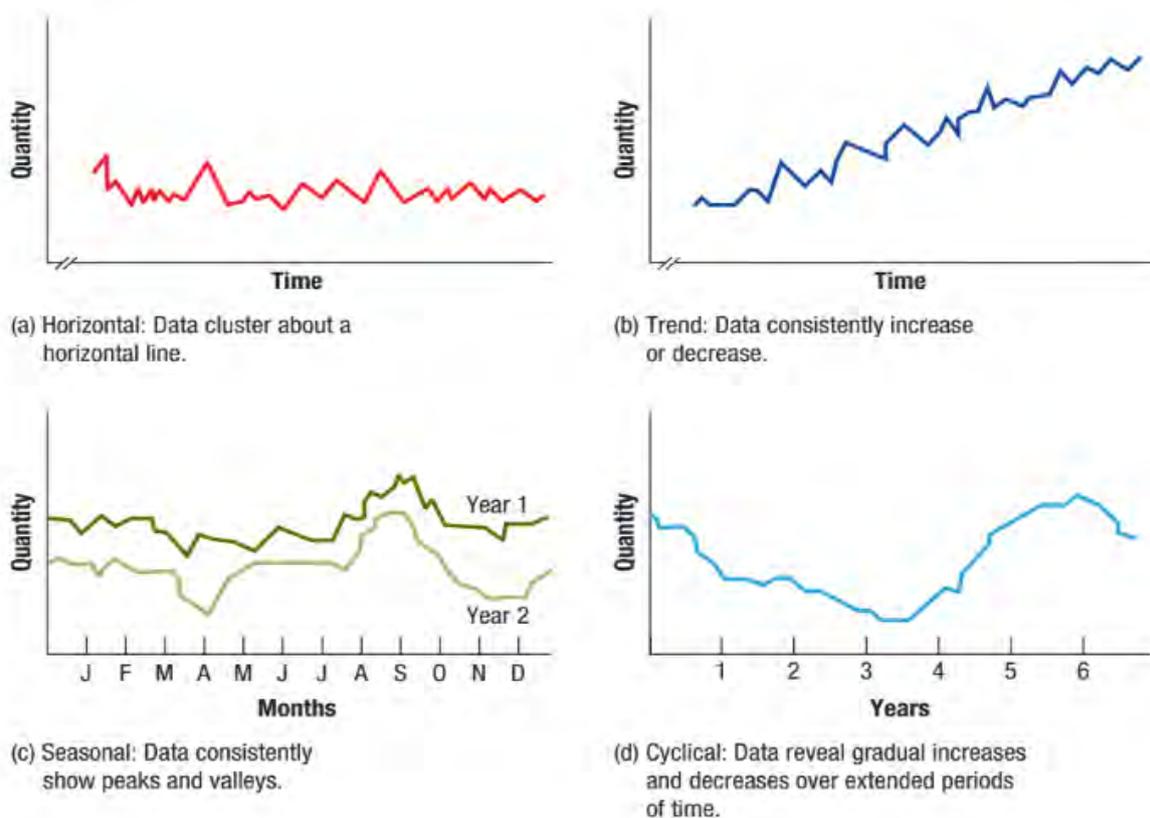


Figura 1. Ejemplos de tendencia de pronosticos  
Tomado de Operations Management, 2018, pp. 278.

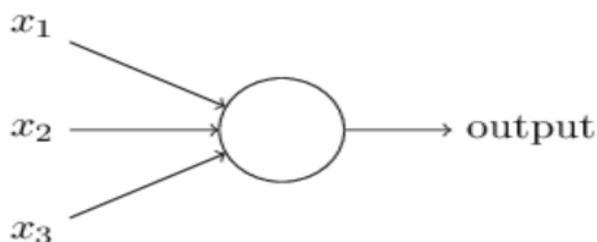
El gráfico (a) de la Figura 1 muestra un patrón horizontal, donde la demanda no varía demasiado y fluctúa en torno a la media. Para este comportamiento, se pueden utilizar métodos como el promedio móvil simple, que toma el promedio de la demanda de los últimos  $n$  periodos, o, si es que se le quiere dar más peso a ciertos periodos, se puede utilizar el periodo móvil ponderado o la suavización exponencial. El gráfico (b) muestra la demanda con tendencia, la cual se caracteriza por presentar un claro comportamiento creciente o decreciente a lo largo del tiempo. Por lo tanto, un método de regresión lineal se podría adecuar bien para estimar la pendiente correspondiente y determinar valores futuros. En el gráfico (c), la demanda es estacional, lo cual implica que existe un patrón que se repite a lo largo de distintos periodos de tiempo; en el ejemplo, se manifiesta a lo largo de los años. Para este caso, se puede usar el método estacional multiplicativo, que toma valores históricos que capturen el comportamiento

estacional y lo reflejan por medio de factores que serán aplicados a cada estación según corresponda. Finalmente, en el gráfico (d) presenta un patrón cíclico, en donde se aprecian variaciones a lo largo de periodos extendidos de tiempo.

### **Fundamentos de Machine Learning**

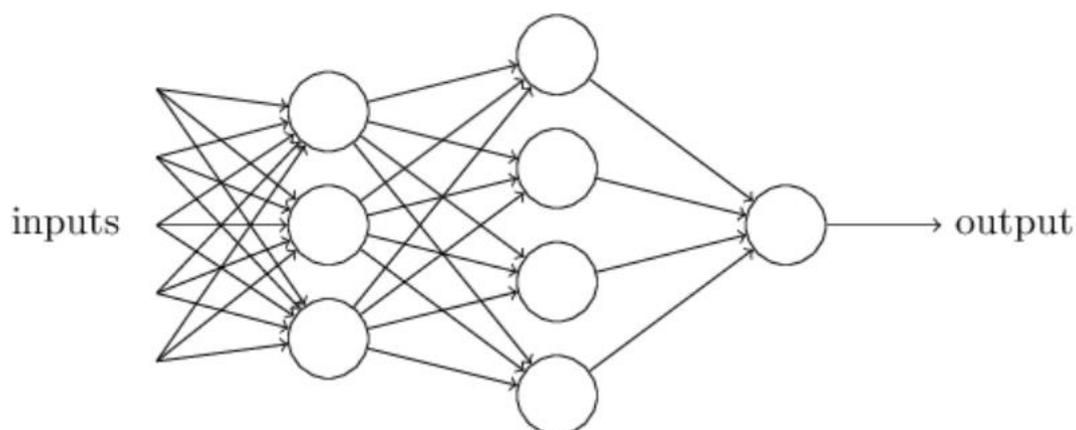
Ahora se explicarán los conceptos que serán útiles para entender la aplicación de estos algoritmos en el cálculo de pronósticos de demanda; para esto, se tomará como referencia el libro Deep Learning de Ian Goodfellow, Yoshua Bengio y Aaron Courville (2016). El objetivo de los algoritmos de Machine Learning es encontrar una función que tome como entrada un grupo de variables y proporcione el valor que se desee estimar. Esta función puede ser de varios tipos; para el caso de pronósticos, se trata de una regresión, ya que el resultado (la demanda) será un número real, a diferencia de una tarea de clasificación, donde se trata de asignar los datos a una categoría representada por un número entero. Para determinar la función que realice la tarea planteada, el modelo se basará en observaciones antiguas, en las cuales el dato que se requiere estimar es conocido. Por ejemplo, si se tienen los datos de la demanda de los últimos 3 años, estos servirán para entrenar al modelo y conseguir la función que logre predicciones acertadas a futuro, cuando el dato de demanda ya no es conocido.

Entre los modelos existentes, las redes neuronales han ganado popularidad por su efectividad y capacidad de personalización. Como explica Nielsen (2015) en su libro Neural Networks and Deep Learning, la estructura básica sobre la cual funcionan las redes neuronales es el perceptrón, también conocido como neurona.



*Figura 2.* Estructura de un perceptrón  
Tomado de Neural Networks and Deep Learning, 2015.

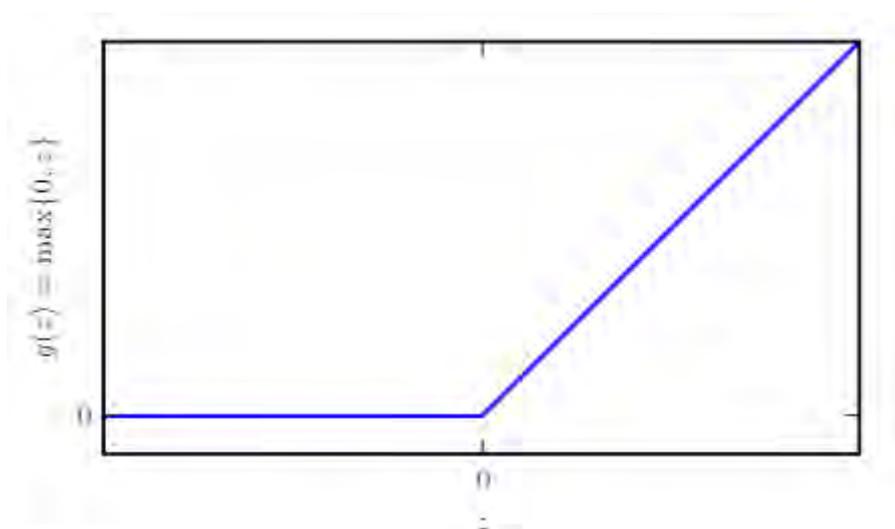
Como se aprecia en la Figura 2, el perceptrón toma tres valores de entrada y devuelve uno de salida. A cada variable de entrada se le asigna un peso  $w$ , los cuales serán utilizados para calcular el output mediante la siguiente fórmula:  $x_1w_1 + x_2w_2 + x_3w_3$ . Para obtener un modelo más complejo, se repetirá esta estructura a lo largo de la red, calculando los outputs de la forma explicada anteriormente. El resultado es el siguiente:



*Figura 3.* Estructura de una red neuronal simple  
Tomado de Neural Networks and Deep Learning, 2015.

Cada arreglo vertical conforma una capa de la red, donde las capas entre los inputs y los outputs se llaman capas ocultas. El número de capas ocultas, así como la cantidad de neuronas dentro de cada una constituye la arquitectura de la red y deberá ser establecida por medio de evaluación. Sin embargo, como explica Goodfellow et al. (2016), una estructura basada únicamente en modelos lineales, como el del ejemplo, implica una limitación significativa, por

lo cual será necesario introducir funciones no lineares para representar adecuadamente las interacciones entre capas. A estas funciones se les llama funciones de activación y serán aplicadas a las salidas de las capas ocultas, y, en algunos casos, en el output final también. La función de activación más utilizada es ReLU (Rectified Linear Unit) y tiene la siguiente forma:



*Figura 4.* Gráfica de la función ReLU  
Tomado de Deep Learning, 2016.

Una vez definida la arquitectura, se entrenará la red de forma que se obtengan los pesos  $w$  óptimos y se minimice el error entre las estimaciones y los valores reales. En la actualidad se pueden llegar a utilizar miles de capas en los modelos de redes neuronales, a lo cual se le llama Deep Learning, o aprendizaje profundo, debido a la gran cantidad de capas ocultas y parámetros disponibles para modificar. Ahora que se han explicado los métodos disponibles para el cálculo de pronósticos, así como la importancia de estos, se presentarán algunas herramientas útiles para el diagnóstico de procesos, las cuales podrían ayudar a determinar la existencia de un problema de pronósticos.

## Diagnóstico de Procesos

Con respecto a los métodos cualitativos, Kaoru Ishikawa (1976) presenta una opción en su Guía para el Control de Calidad: el diagrama causa-efecto, también conocido como espina de Ishikawa. Originalmente, esta herramienta fue concebida para apoyar a determinar los factores que generan dispersión en la calidad de procesos de manufactura; sin embargo, también puede ser aplicada al diagnóstico de procesos de forma más general. Se empieza por elegir cuál es la característica que se quiera mejorar, la cual representa un problema en el proceso. Luego, se coloca al lado derecho del diagrama, acompañada de la rama principal, llamada así pues será en torno a este problema que se realizará el análisis. Ahora se deben establecer los factores que, a nivel general, estén vinculados al inconveniente en cuestión, y estos se colocan en ramas que se desprenden de la rama principal, indicando una relación de causalidad. Finalmente, se van añadiendo elementos causales a las ramas de forma iterativa y llegando hasta el nivel de detalle que sea necesario, hasta que se reflejen por completo las causas vinculadas al problema principal. La siguiente figura muestra un ejemplo del resultado final del ejercicio descrito:

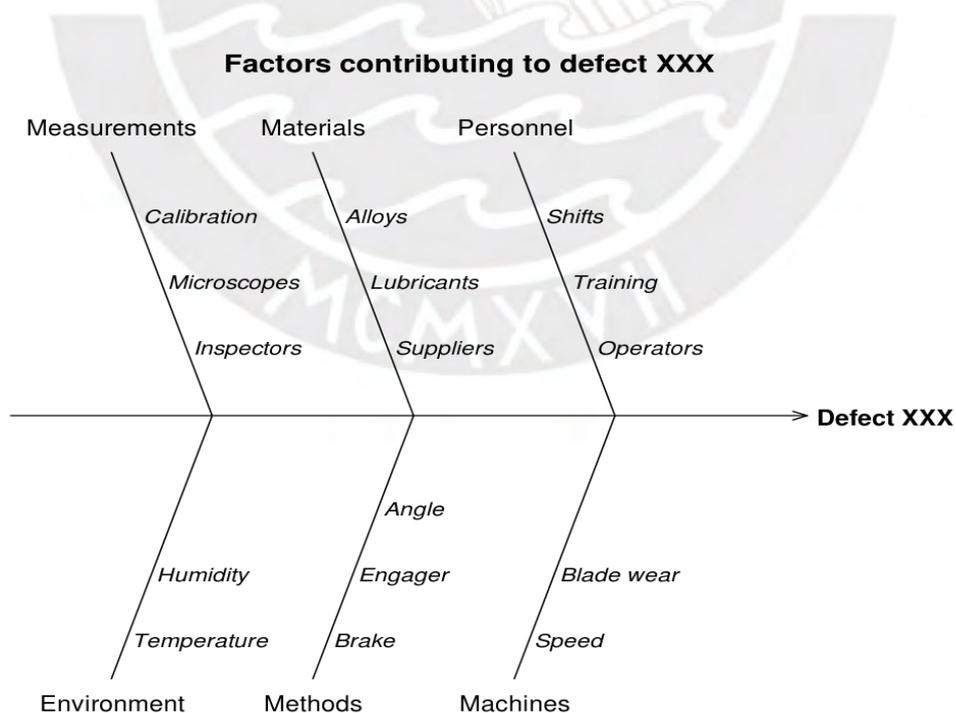
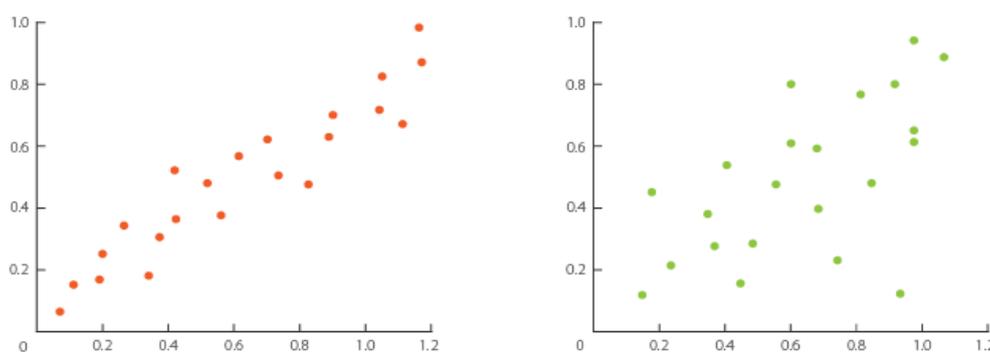


Figura 5. Ejemplo de Espina de Ishikawa  
Tomado de Guide to Quality Control, 1976.

El método demanda que la elaboración del diagrama se realice con todos los miembros del equipo involucrado en el proceso, por medio de una lluvia de ideas, de forma que el resultado sea completo. Del mismo modo, una vez obtenido el esquema, se procederá a analizarlo y a decidir cuáles son las causas más probables al problema que se quiera solucionar, y se elaborará un plan de acción orientado a lidiar con los factores identificados (Ishikawa, 1976).

Cuando tanto las causas elegidas como el problema en estudio tienen carácter cuantitativo, una alternativa es aplicar un análisis de regresión para comprobar la relación lineal entre ambos factores. Por lo tanto, se parte desde el resultado del diagnóstico cualitativo y se define al factor escogido como la variable independiente de la regresión. Hacer una gráfica de dispersión de ambas variables puede ayudar a visualizar el grado de correlación, como en el siguiente ejemplo:



*Figura 6.* Diferencias entre distintos niveles de correlación  
Tomado de The Data Visualisation Catalogue. Recuperado de:  
<https://datavizcatalogue.com/methods/scatterplot.html>

El gráfico de la izquierda muestra una correlación lineal más fuerte que el de la derecha, donde los puntos son más dispersos, asimismo, esta correlación es positiva al presentar un comportamiento creciente. Numéricamente, se puede utilizar el coeficiente de correlación de Pearson para determinar el nivel de asociación lineal, utilizando la siguiente fórmula:

$$\rho_{x,y} = R = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{n \sum_i^n (x_i y_i) - \sum_i^n x_i \sum_i^n y_i}{\sqrt{n \sum_i^n x_i^2 - (\sum_i^n x_i)^2} \sqrt{n \sum_i^n y_i^2 - (\sum_i^n y_i)^2}}$$

Donde:

- $n$  = número de muestras
- $x_i$  = valores de la variable independiente
- $y_i$  = valores de la variable dependiente
- $\rho_{x,y}$  = coeficiente de correlación de Pearson

El valor del coeficiente puede variar entre -1 y 1, donde los valores negativos indican que la relación es decreciente. Asimismo, si es que el valor obtenido es muy cercano a 0, se podría concluir que no hay correlación lineal entre ambas variables; no obstante, podría existir otro tipo de relación entre estas. De todas formas, este indicador servirá de apoyo junto a los métodos cualitativos y ayudará a entender mejor el comportamiento de las variables de análisis.

## Contenido de la investigación

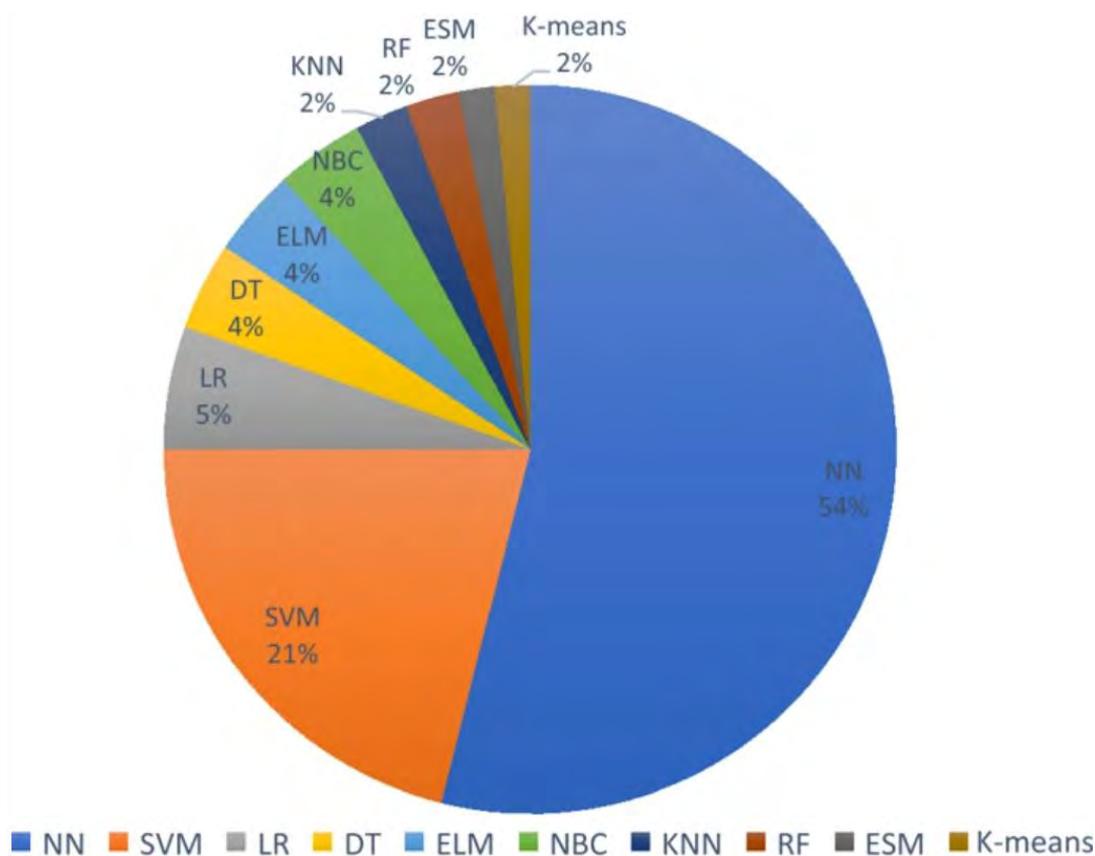
A continuación, se presentará una serie de artículos y publicaciones académicas que muestran la aplicación de algoritmos de machine learning en pronósticos y Supply Chain.

### **1. A systematic review of the research trends of machine learning in supply chain management, Du Ni, Zhi Xiao y Ming K. Lim (2019)**

En este artículo se aborda una investigación en la cual se revisan diversos trabajos académicos de aplicaciones de machine learning en actividades de gestión de cadena de valor. Para la investigación en mención, los autores buscaron en múltiples bases de datos (Emerald Insight, IEEE Xplore, Scopus, Science Direct y Springer) por publicaciones con “machine learning” y “supply chain management” como palabras clave, para finalmente obtener una muestra de 123 artículos. Asimismo, en la introducción se detallan las ventajas de este tipo de soluciones, en comparación con los métodos tradicionales, como por ejemplo la posibilidad de capturar patrones no lineales, la posibilidad de lidiar con datos no estructurados y un mejor desempeño predictivo; sin embargo, se especifica que tan solo el 15% de empresas utiliza ML para tratar alguna actividad de supply chain (Ni et al., 2019).

De las 123 publicaciones revisadas en el artículo en cuestión, más del 40% se concentra en China y Estados Unidos, lo cual está relacionado con el alto nivel de industrialización de estos países; en contraste, solo una de estas corresponde a un país latinoamericano (Chile). Por otro lado, se utilizan diversos métodos para obtener la data necesaria para entrenar los modelos de las investigaciones, donde la información generada mediante simulaciones es la opción más popular, presente en el 50% de los artículos, y seguida por el uso de datos históricos, con el 33%. Como se puede observar en la Figura 7, los algoritmos más empleados fueron las redes

neuronales (54%) y las support vector machines (21%), donde las primeras destacan por su flexibilidad y adaptabilidad a los datos; mientras que las segundas se caracterizan por una gran interpretabilidad y generalización (Ni et al., 2019). Finalmente, se especifica que, de la variedad de actividades comprendidas por Supply Chain, la estimación de la demanda es el tema con mayor presencia en las investigaciones que fueron revisadas.



*Figura 7.* Distribución de los algoritmos encontrados en las investigaciones Tomado A systematic review of the research trends of machine learning in supply chain management, 2019.

## **2. A neural network solution for forecasting labor demand of drop-in peer tutoring centers with long planning horizons, Rick Brattin, Randall S. Sexton, Wenqiang Yin y Brittaney Wheatley (2019)**

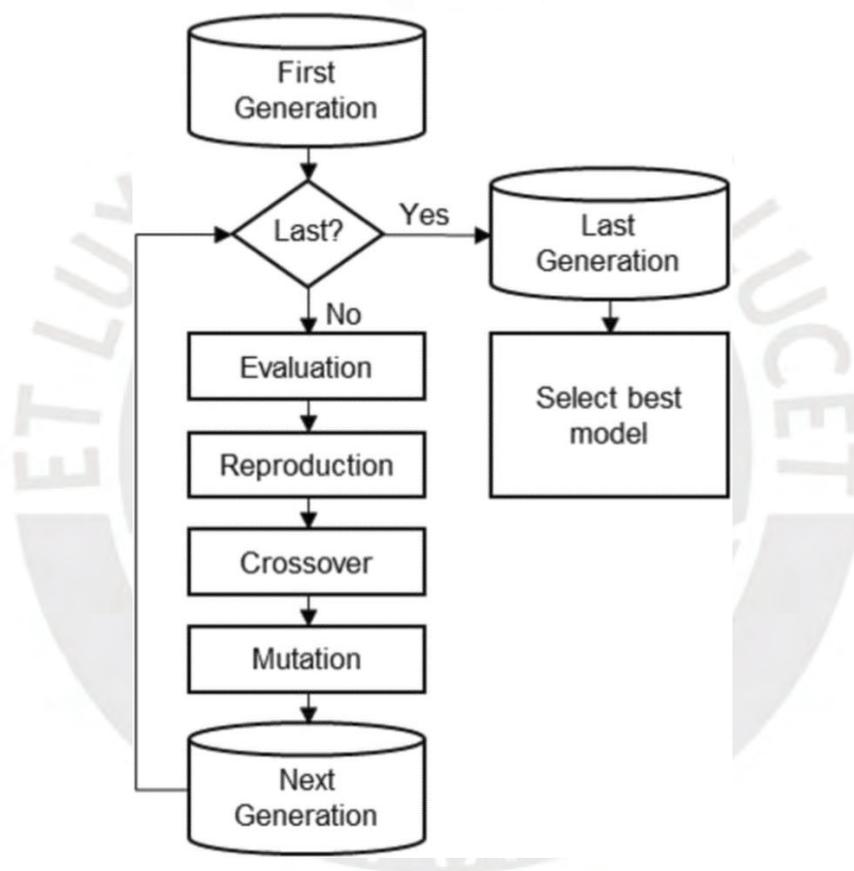
En este artículo se presenta una solución de redes neuronales para pronosticar la demanda en centros de tutoría estudiantil, con lo cual se generaría la asignación de personal adecuada. Se empieza mencionando la importancia de lograr una buena estimación de la demanda, pues el

éxito de este tipo de servicios depende de que se cuente con la suficiente cantidad de tutores para orientar a los estudiantes interesados, mientras que un exceso de oferta resultaría en un gasto excesivo. Asimismo, se especifica que la planificación debe realizarse a largo plazo dado que, por el nivel de especialización requerido para el personal, hallar un reemplazo en un corto plazo es inviable. Con respecto a la base de datos disponible para el entrenamiento, se utilizó información histórica extraída del sistema de reserva del centro de tutoría de una universidad con 24,000 estudiantes, tomando datos desde el 2010 hasta el 2018. Con esto, se obtuvieron más de 72,000 registros con información como curso, fecha, hora de entrada y hora de salida.

En relación al método utilizado, se compararon varios modelos y sus respectivos errores medios cuadráticos. Se plantearon dos modelos lineares para tomar como base, un promedio móvil tomando datos de cada 5 semanas para el último año, y una regresión lineal. Por otro lado, se elaboraron cuatro redes neuronales con distintas variables de input para cada una. La primera red toma únicamente las variables base, las mismas utilizadas para la regresión lineal; en la segunda se añade el promedio móvil como una variable adicional; la tercera red incorpora los resultados de la regresión lineal como otro input y; la cuarta red introduce los resultados del promedio móvil y los de la regresión.

En cuanto al método de entrenamiento elegido, a diferencia de lo que se acostumbra, no se utilizó el algoritmo de backpropagation para hallar los parámetros que definirían a la red, sino que se empleó un enfoque de algoritmos genéticos, que se conoce como Neural network simultaneous optimization algorithm (NNSOA). Este último, no solo ayudaría a definir parámetros como los pesos de la red, sino que también encuentra la cantidad óptima de nodos por capa, basándose en procesos de evolución y selección natural, por medio de múltiples iteraciones. Este método funciona al ir incrementando la cantidad de nodos en la red, de forma

que se determina el valor óptimo una vez que el nodo añadido ya no mejore el desempeño de esta. Por otro lado, para los pesos se generan soluciones aleatorias de forma iterativa y se evalúan sus desempeños, los cuales determinarán la probabilidad de que en posteriores generaciones estas estén presentes. Estas soluciones son la base de cada generación pues estas son creadas por medio de la combinación de las soluciones anteriores, por lo cual en cada iteración estas presentarán pesos cuyo error sea menor. En la Figura 8 se muestra el flujo que explica este proceso.



*Figura 8.* Funcionamiento del algoritmo NNSOA

Tomado de A neural network solution for forecasting labor demand of drop-in peer tutoring centers with long planning horizons, 2019.

Finalmente, los resultados indicaron que los cuatro modelos que utilizaron redes neuronales tuvieron menor porcentaje de error que los modelos base, reduciéndolo desde 1% hasta 10%, logrando así un mejor ajuste a la demanda y una mayor capacidad predictiva. De estos, la red

neuronal con mejores resultados fue la segunda, que combina las variables base con los resultados del promedio móvil.

### **3. Prediction of air travel demand using a hybrid artificial neural network (ANN) with Bat and Firefly algorithms: a case study, Ali Mostafaeipour, Alireza Goli y Mojtaba Qolipour (2018)**

El uso de algoritmos inspirados en la naturaleza no es algo poco común, por lo que también se muestra su efectividad en un caso de estudio para determinar la demanda de vuelos en una aerolínea iraní. En este artículo se presentan dos algoritmos meta heurísticos: Firefly y Bat, y se compara la efectividad de ambos al ser aplicados para determinar los mejores pesos dentro de una red neuronal. Ambos tienen como fundamento la búsqueda aleatoria y buscan aumentar la probabilidad de encontrar un mínimo absoluto con respecto a la función de error de estimación; los métodos tradicionales suelen “atorarse” en puntos de silla que significan mínimos locales, por lo cual su desempeño puede no ser óptimo. Al comparar los resultados de una regresión lineal, una red neuronal estándar, una híbrida con el algoritmo Bat y una híbrida con el Firefly, se obtienen los siguientes resultados:

Tabla 1  
*Resumen de resultados al comparar ANNs y regresión*

Criterios	Regression	Artificial neural network		
		ANN + FA	ANN + BA	ANN
MAE	391.24	1.36	9.34	347.29
MSE	9975.86	34.97	475.1	9425.3
MAPE	9.11	0.641	3.11	7.36
MdAPE	0.73	0.10	0.21	0.61
$R^2$	0.2116	0.9554	0.9016	0.2339

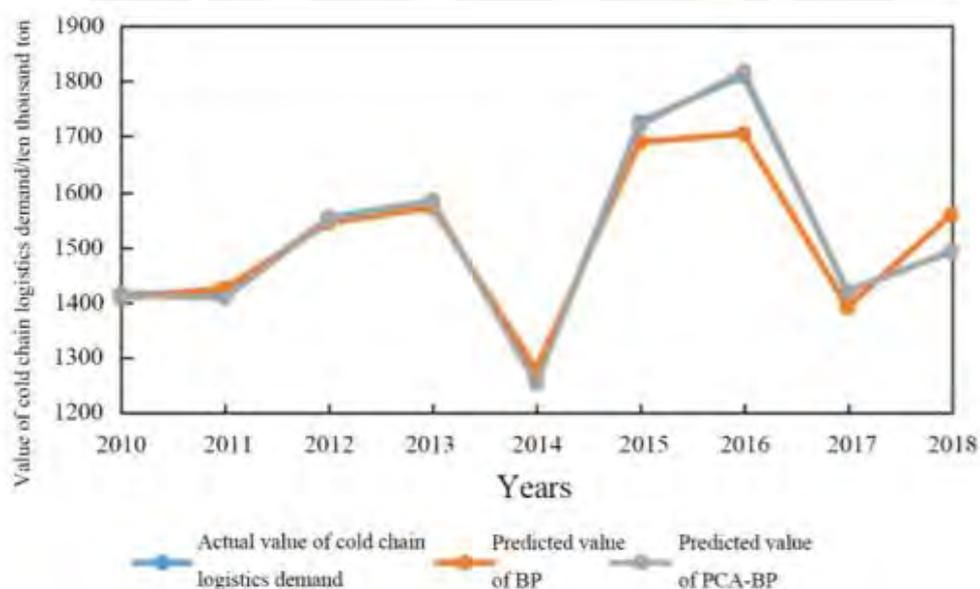
Nota. Tomado de Prediction of air travel demand using a hybrid artificial neural network (ANN) with Bat and Firefly algorithms: a case study, Ali Mostafaeipour, Alireza Goli y Mojtaba Qolipour (2018)

Se puede observar que el nivel de ajuste, determinado por el  $R^2$ , es significativamente mayor para los modelos que utilizaron los algoritmos mencionados; mientras que el error, determinado por los indicadores MSE y MAE, es mucho menor. Además, se obtiene que el algoritmo Firefly

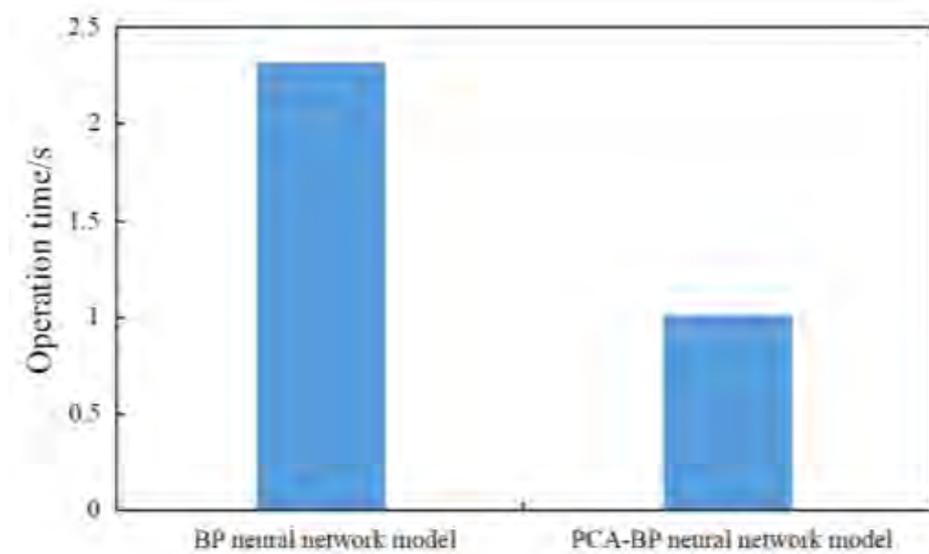
obtiene los mejores resultados de los cuatro modelos mostrados, lo cual corrobora su efectividad para optimizar redes neuronales.

#### 4. Urban cold-chain logistics demand predicting model based on improved neural network model, Ying Chen, Qiuming Wu y Lijin Shao (2020)

En esta publicación se menciona la aplicación de PCA (Principal Component Analysis), en conjunto con redes neuronales, en la demanda de alimentos frescos y su influencia para la logística de cold chain. La importancia de la determinación de demanda para esta industria es importante pues este tipo de alimento es perecible y se puede incurrir en pérdidas si es que se planifican inadecuadamente los recursos necesarios para manejar una cold chain. El artículo se centra en la implementación de PCA para reducir el número de variables necesarias para la implementación de la red neuronal. Esto se logra realizando un análisis de contribución de varianza, tomando como variables finales aquellas que presenten mayor contribución. En el caso estudiado solo se utilizan la mitad de las variables originales, que aportan más del 98% de la varianza de los datos disponibles. Los resultados obtenidos son los siguientes:



*Figura 9.* Gráfico de líneas de valores reales y predicciones con BP y PCA-BP  
Tomado de Urban cold-chain logistics demand predicting model based on improved neural network model, 2020.



*Figura 10.* Diferencias de tiempo operacional entre BP y PCA-BP Tomado de Urban cold-chain logistics demand predicting model based on improved neural network model, 2020.

Las Figura 9 muestran que las predicciones realizadas por el modelo que utilizó PCA fueron más cercanas a la demanda real en todos los años evaluados. Asimismo, en la Figura 10 se puede apreciar que el tiempo de ejecución del modelo regular es mayor, donde este último tardó más del doble que la red con PCA. Estas mejoras se deben a que, al utilizar un menor número de variables, se disminuyen los cálculos requeridos y se aumenta la eficiencia; por otro lado, dado que no se están incluyendo las variables menos significativas, se reducen efectos perjudiciales para el desempeño del modelo.

### **5. Supply–Demand Prediction for Agile Manufacturing with Deep Neural Network, Rong Wen y Wenjing Yan (2019)**

A diferencia de las propuestas anteriores, en este artículo se propone un modelo predictivo basado en una red neuronal simple, pero que utiliza múltiples tipos de dato para lograr reducir la brecha entre oferta y demanda para la industria manufacturera. En la arquitectura planteada, se implementan capas para realizar procesamiento y concatenación de datos (Wen y Yan, 2019). Por ejemplo, información como el código de almacén y la fecha son combinados para

crear una variable espacio temporal, que exprese relaciones de cercanía en ambos aspectos. Asimismo, se habilita un bloque de posibles factores externos que podrían afectar la predicción deseada, acá se podrá incluir información climática, datos del mercado y eventos recientes. Finalmente, los resultados indican una mejoría con respecto a los modelos basados en series de tiempo, como ARIMA, promedio móvil y suavización exponencial; también se obtienen predicciones más precisas que un algoritmo de redes neuronales más complicado, como LSTM (Long Short Term Memory Network), concluyendo la efectividad del procesamiento de datos para el rendimiento del modelo.

El flujo de procesamiento de la red aplicada se muestra a continuación:

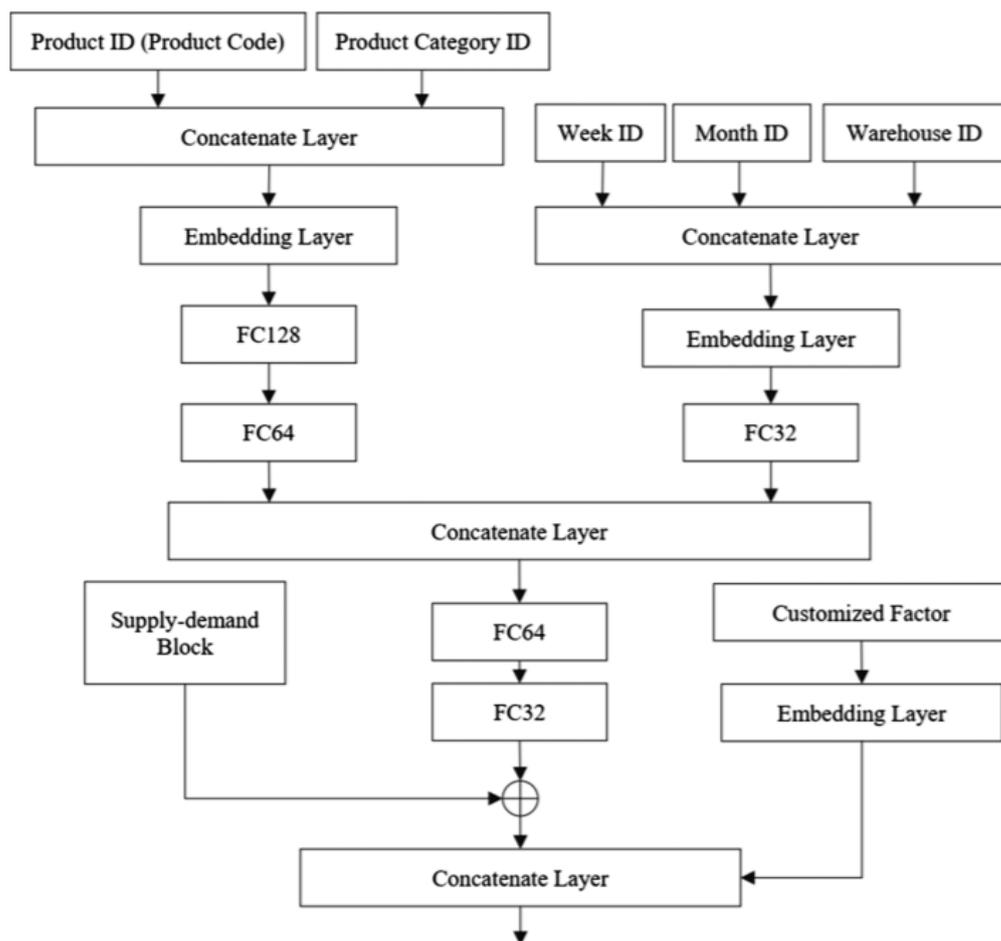


Figura 11. Diagrama de flujo de la arquitectura de una red neuronal que utiliza Feature Engineering Tomado de Supply–Demand Prediction for Agile Manufacturing with Deep Neural Network, 2019.

## 6. Supply Chain Demand Forecasting; A Comparison of Machine Learning Techniques and Traditional Methods, J. Shahrabi, S. S. Mousavi y M. Heydar (2009)

De forma similar, en una investigación aplicada a estimar la demanda en una compañía de carros iraní, se obtuvieron resultados positivos al utilizar redes neuronales. En este caso, el patrón del caso en estudio es de tipo estacional con tendencia, por lo cual se requerirá una solución que capture tendencias no lineales. En el estudio se presentan dos soluciones basadas en machine learning: Support Vector Machine y una red neuronal. En la primera propuesta, se debe elegir una función de kernel, que está basada en productos de vectores, que se adecúe mejor para el caso, entre las cuales está la función lineal, logarítmica, radial, sigmoïdal, etc. Este paso es fundamental para obtener resultados fiables y que el modelo pueda adoptar el patrón necesario. En el caso de la red neuronal, se optó por una solución que contaba con únicamente una capa oculta; sin embargo, a pesar de su simplicidad, esta solución es capaz de recoger los patrones no lineales sin la necesidad de establecer explícitamente una función, como en el caso de SVM.

Para comparar los modelos, se empleó el MAPE, que representa el error porcentual medio entre las predicciones y el valor real. En la Tabla 2, se muestra la comparación del desempeño de cada propuesta:

Tabla 2

*Resumen de resultados al comparar SVM, ANN y diversos modelos clásicos*

Forecasting techniques	Testing data set (rank)	Training data set (rank)
Moving average	167.753 (2)	180.604 (3)
Exponential smoothing	167.797 (3)	179.791 (4)
Exponential smoothing with trend	170.081 (4)	187.051 (5)
Support vector machines	171.282 (5)	178.540 (2)
Artificial neural network	163.310 (1)	169.141 (1)

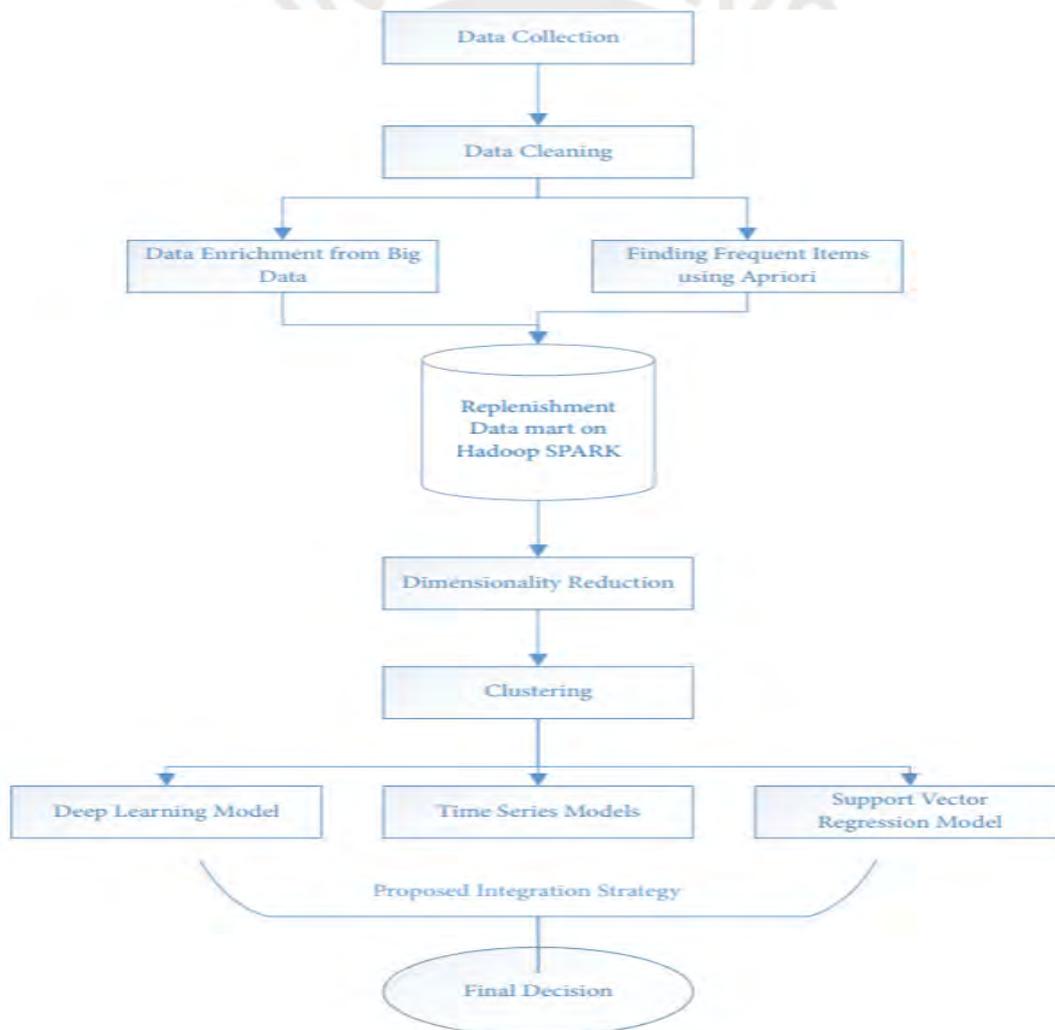
*Nota.* Tomado de Supply Chain Demand Forecasting; A Comparison of Machine Learning Techniques and Traditional Methods, 2009.

Al analizar el error en la data de entrenamiento, se observa que ambas soluciones de machine learning tienen los mejores resultados; no obstante, al evaluar con información nueva (testing data set), la SVM pasa a tener el mayor error, lo cual indica que se falló al generalizar para data nueva. Cabe resaltar que los resultados obtenidos en el testing data set son más relevantes para decidir si un algoritmo es mejor que otro, ya que estos reflejan cómo se comportaría el modelo con datos nuevos para este, lo cual es la situación a la que se debe enfrentar un modelo al ponerlo en producción. Finalmente, para ambos data sets, el mejor modelo fue la red neuronal.

### **7. An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain, Zeynep Hilal Kilimci; A Okay Akyuz; Uysal, Mitat; Akyokus, Selim; Uysal, M Ozan; et al. (2019)**

En una publicación reciente, se propone una alternativa novedosa para combinar los resultados de diversos modelos predictivos, para calcular el pronóstico de demanda de un Retail llamado SOK Market. Cuando un modelo surge de la combinación de otros, se habla de un ensamble, y hay muchas formas de determinar el resultado final de la predicción. Uno es hallar los errores de cada modelo en un periodo determinado y, para calcular la predicción del periodo siguiente, se utiliza únicamente el valor de aquel modelo con el mejor error, por lo que se irán alternando los algoritmos utilizados conforme avance la serie de tiempo. La metodología que se presenta en el artículo implica la ponderación de las predicciones arrojadas por cada algoritmo, de forma que se realice de forma “democrática”. Para hallar los pesos correspondientes, se aplica un concepto similar al utilizado en la primera metodología; es decir, se calculan los errores de un periodo, pero, en vez de tomar el valor del mejor modelo directamente, se determinan pesos en función de dichos errores, los cuales multiplicaran a la predicción de cada algoritmo favoreciendo más a aquellos con mejor desempeño.

Para el caso en cuestión se eligieron once estimadores base sobre los cuales se implementaría el ensamble, entre los cuales se encuentran nueve modelos de series de tiempo (promedio móvil, suavización exponencial, ARIMA, regresiones, etc.), un Support Vector Machine y un modelo de Deep Learning (red neuronal). Por otro lado, debido al alto volumen de datos de la base disponible (875 millones de registros y 155 variables), resulta poco eficiente el entrenamiento del modelo de redes neuronales, por lo cual es necesario aplicar reducción de dimensiones (PCA) para disminuir el número de variables y clustering para agrupar los productos en categorías, con el objetivo de no entrenar un modelo por cada producto, sino que solo para cada categoría. El flujo empleado se muestra a continuación:



*Figura 12.* Diagrama de flujo de una arquitectura que combina Deep Learning, SVM y series de tiempo

Tomado de An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain, 2019.

Finalmente, se plantearon los escenarios S1, S2 y SD para comparar sus respectivos resultados, donde: S1 utiliza el método de ensamble tradicional y no incluye al modelo de Deep Learning, S2 utiliza el método de ensamble propuesto y tampoco incluye al modelo de Deep Learning y SD utiliza el método propuesto y sí incluye al modelo de Deep Learning. Al comparar estos escenarios, tanto S2 como SD presentan menor error que S1, corroborando la efectividad de la estrategia de ensamble propuesta; asimismo, SD tiene mejor desempeño que S2, con lo que se verifica la capacidad predictiva de las redes neuronales para capturar patrones irregulares como los de demanda.

#### **8. Forecasting of sales by using fusion of Machine Learning techniques, Mohit Gurnani, Yogesh Korkey, Prachi Shahz, Sandeep Udmalex, Vijay Sambhe y Sunil Bhirudk (2017)**

En este artículo también se utiliza la combinación de diversos modelos para lograr mejores resultados. En este caso se trata de predecir el pronóstico de ventas de una empresa farmacéutica llamada Rossmann. En la investigación se plantean y comparan diversas alternativas, entre las cuales se encuentran modelos individuales como ARIMA, redes neuronales, XGBoost y SVM; el enfoque híbrido, que implica descomponer la predicción en un patrón lineal y un patrón no lineal (el residuo); y una descomposición aplicando STL, un método en el cual el pronóstico se divide en tres componentes: la estacionalidad, la tendencia y el residuo. Tanto el enfoque híbrido como la descomposición con STL combinan diversos modelos y los integran por medio de la suma de estos. Para lograr esto, se prueban diversos modelos para cada componente; en el caso del híbrido, se toma ARIMA como el algoritmo principal para determinar el patrón lineal, y se prueba con distintos modelos para hallar la diferencia entre la predicción de ARIMA y los valores reales. En cuanto a STL, SNaive presenta menor error al predecir el componente estacional de las ventas, ARIMA obtuvo los

mejores resultados para el componente de tendencia y XGBoost fue el mejor modelo para obtener los residuos. La consolidación de los resultados se muestra en la Tabla 3:

Tabla 3

*Resumen de resultados al comparar combinaciones de ARIMA, ARNN, XGBoost y SVM*

Performance Measure	MAE	RMSE
ARIMA	481	771
ARNN	413.4	565.5
XGBoost	346.4	670
SVM	380.6	666.7
Hybrid ARIMA-ARNN	372.8	530.4
Hybrid ARIMA-XGBoost	357	540.5
Hybrid ARIMA-SVM	407.4	610
STL Decomposition	328.6	426.4

*Nota.* Tomado de Forecasting of sales by using fusion of Machine Learning techniques, 2017.

Se observa que, de los modelos individuales, la red neuronal obtiene los mejores resultados debido a su capacidad para determinar patrones complejos; sin embargo, al combinar los modelos se obtiene mejor desempeño. En todos los casos, el enfoque híbrido reduce el error de predicción, siendo la combinación ARIMA-ARNN la que produce resultados más precisos. Aun así, la mejor integración de modelos viene dada por la descomposición STL, en donde se aplican modelos aplicados a un componente en específico y representa la solución más robusta para la predicción deseada.

### **9. Analysis on Machine Learning Algorithms and Neural Networks for Demand Forecasting of Anti-Aircraft Missile Spare Parts, Nikita Pawar y Bhavana Tiple (2019)**

En esta publicación se presenta una alternativa para el pronóstico de demanda de partes de misiles antiaéreos. Para este caso se cuenta con registros de 10,000 misiles correspondientes a la base de Vietnam War Bombing Operations. La metodología empleada implica el desarrollo

de diversos modelos para comparar su capacidad predictiva, pero también se aplica Feature Engineering, donde se generan nuevas variables que mejoren el desempeño de dichos modelos. Uno de los modelos elegidos es XGBoost, que es de por sí un tipo de ensamble homogéneo, pues basa su funcionamiento en la creación iterativa de árboles de decisión que, en conjunto, logran muy buenos resultados. A diferencia de la propuesta del artículo anterior, este tipo de ensamble solo considera modelos del mismo tipo. En la Tabla 4 se presenta el resumen obtenido para la precisión de predicciones.

Tabla 4

*Resumen de resultados al comparar XGBoost, MLP y diversos modelos de machine learning*

<b>Algorithms</b>	<b>Accuracy with existing features</b>	<b>Accuracy with new features</b>
SVM	76	93
Random Forest	85	93
Logistic Regression	79	90
Decision Tree	83	90
MLP	93	97
XGBoost	87	98

*Nota.* Tomado de Analysis on Machine Learning Algorithms and Neural Networks for Demand Forecasting of Anti-Aircraft Missile Spare Parts, 2019.

Se puede observar que, al añadir nuevas variables significativas, el rendimiento de todos los algoritmos aumenta. Asimismo, si es que se utiliza la base sin aplicar Feature Engineering, la red neuronal (MLP o Multi Layer Perceptron) supera al resto de modelos; sin embargo, con las nuevas variables el algoritmo XGBoost logra resultados ligeramente superiores.

#### **10. Analysis of book sales prediction at Amazon marketplace in India: a machine learning approach, Satyendra Kumar Sharma, Swapnajit Chakraborti y Tanaya Jha (2019)**

El uso de algoritmos de Machine Learning puede ayudar a encontrar las variables más influyentes para la demanda. Una publicación se enfoca en esto, al aplicar modelos de regresión

y redes neuronales para predecir la demanda de libros en Amazon, y luego obtener retroalimentación sobre el desempeño de cada variable sobre la predicción. La idea es que, la significancia de una variable en los diversos modelos se relaciona con la importancia de esta para propósitos del negocio y las ventas. El nivel de importancia de una variable para el modelo se puede determinar de muchas formas; para una regresión lineal, por ejemplo, aportarán más al modelo aquellos factores que cuenten con un coeficiente mayor. Del mismo modo, en un modelo de árbol de decisión, las variables más significativas son las que aparecen en más ramas de decisión del árbol definido. Identificar estas variables es valioso para tomar decisiones adecuadas y entender mejor el comportamiento de la demanda.

En esta investigación, también se evaluó la interacción entre las diversas variables y se incluyeron resultados de sentiment analysis de los comentarios de la página web. Entre las conclusiones halladas, se encuentra que la significancia de la Tasa de Descuento es mayor que la de Monto de Descuento; sin embargo, cuando la Cantidad de Comentarios es alta, la variable Tasa de Descuento se vuelve aún más importante, lo cual indica una interacción positiva entre estas. Esto puede ayudar a determinar que es efectivo aplicar descuentos a artículos con alto volumen de comentarios o reviews. Finalmente, se resalta que, en adición a la capacidad predictiva de estos algoritmos, estos pueden brindar intuiciones derivadas directamente de los patrones hallados en la data, lo cual puede beneficiar significativamente a los objetivos estratégicos de una empresa.

## Discusión

En la sección previa se revisaron diez artículos relacionados al tema de investigación propuesto, cada uno exponiendo diversos enfoques que se le podía dar a una solución que involucre machine learning y redes neuronales en un problema de Supply Chain. A continuación, se contrastará el contenido expuesto en estas publicaciones y se buscará presentar un análisis detallado de las similitudes y diferencias en estos enfoques.

Para empezar, la investigación 1 no presenta una solución como tal, sino que es una revisión de literatura realizada por el autor, lo cual brinda una idea general de cómo se ha ido abordando este tema en el ámbito académico. Asimismo, en esta se indica que el nivel de investigación de este ámbito en Latinoamérica es muy bajo en comparación a otros países. Al realizar este trabajo de investigación, se corroboró lo expuesto por este autor, ya que la mayoría de propuestas encontradas se basaban en el uso de redes neuronales, y ninguna de estos artículos fueron publicados por revistas o universidades latinoamericanas.

Continuando con la discusión, los artículos 2 y 3 presentan una similitud en cuanto a la forma en la que se optimiza la red neuronal utilizada. En el 2 se presenta una optimización por medio de algoritmos genéticos, mientras que la 3 contrasta el uso del algoritmo Firefly y el algoritmo Bat. Estos tres métodos se caracterizan por tomar como referencia procesos de la naturaleza para resolver un problema que suele ocurrir en Deep Learning, que es encontrar los parámetros óptimos de la red. Estos enfoques son distintos al método tradicional, que es el algoritmo de backpropagation y que está basado en el cálculo diferencial para ir minimizando el error iterativamente. En ambos artículos, la aplicación de estas técnicas novedosas consiguió una mejora en los resultados y predicciones en cuestión, lo cual es una muestra de la efectividad de

estos. Sin embargo, una desventaja es que la mayoría de librerías y frameworks de Deep Learning tienen ya implementado el backpropagation, por lo cual, de querer utilizar los algoritmos presentados, sería necesario desarrollarlos manualmente, con lo cual se sacrifica practicidad y optimización computacional. Además, no es claro si es que estos resultados positivos pueden ser generalizables para un rango más amplio de problemas, a diferencia del método tradicional, cuyo desempeño es consistentemente satisfactorio. Por esto, debería evaluarse si es significativa la mejora obtenida con los algoritmos no tradicionales como para asumir la implementación de estos.

En cuanto a la publicación 4, en esta el tema central evaluado fue la utilización de PCA para lograr reducción dimensional. Esta técnica también es aplicada en la solución propuesta en el artículo 7, en el cual, debido a la gran magnitud del dataset, no era factible trabajar con las 155 variables originales. Según lo explicado en el artículo 4, aplicar PCA puede generar beneficios en diversos aspectos, como es la reducción del tiempo computacional necesario para correr los diversos modelos, y la reducción de varianza, al reducir el número de predictores. Para el caso de la solución en la publicación 7, se aprovecharon estos beneficios, dado que se buscaba probar diversos modelos y compararlos entre sí, lo que implica que se evalúen y ejecuten los modelos numerosas veces. Además, estos modelos incluían redes neuronales, que requiere mayor tiempo y recursos que otros algoritmos. Finalmente, gracias a que PCA obtiene componentes buscando maximizar la representación de las variables, no se pierde poder predictivo al utilizarlo. Por lo tanto, es recomendable hacer uso de PCA en los casos descritos y se puede obtener resultados positivos como los expuestos en 4 y 7.

Por otro lado, en 5 y 9 también se plantean técnicas para manipular variables, pero no con el enfoque de reducción de dimensionalidad, sino que se crean nuevas a partir de otras, o se

transforman con el fin de mejorar los resultados del modelo. En el artículo 5 se plantea una arquitectura donde se generan variables a partir de información del producto, de información cronológica (mes, día) y de información geográfica. Posterior a este procesamiento, se aplicó un modelo sencillo de redes neuronales que obtuvo mejores resultados que enfoques de series de tiempo como ARIMA y redes recursivas complejas como LSTM. De forma similar, en 9 se prueban diversos modelos antes y después de añadir nuevas variables y, en todos los casos, el performance del modelo aumentó. Asimismo, con las nuevas variables el modelo XGBoost tuvo menor error que la red neuronal, a pesar de que esta última es más compleja. Por lo tanto, en algunos casos puede resultar más conveniente dedicarle mayor tiempo en la etapa de preprocesamiento y Feature Engineering antes que utilizar un modelo complejo. Esto puede contradecir el enfoque del resto de investigaciones, donde se prioriza el modelo como tal y no la creación de variables que podrían tener un efecto mayor que la optimización del algoritmo.

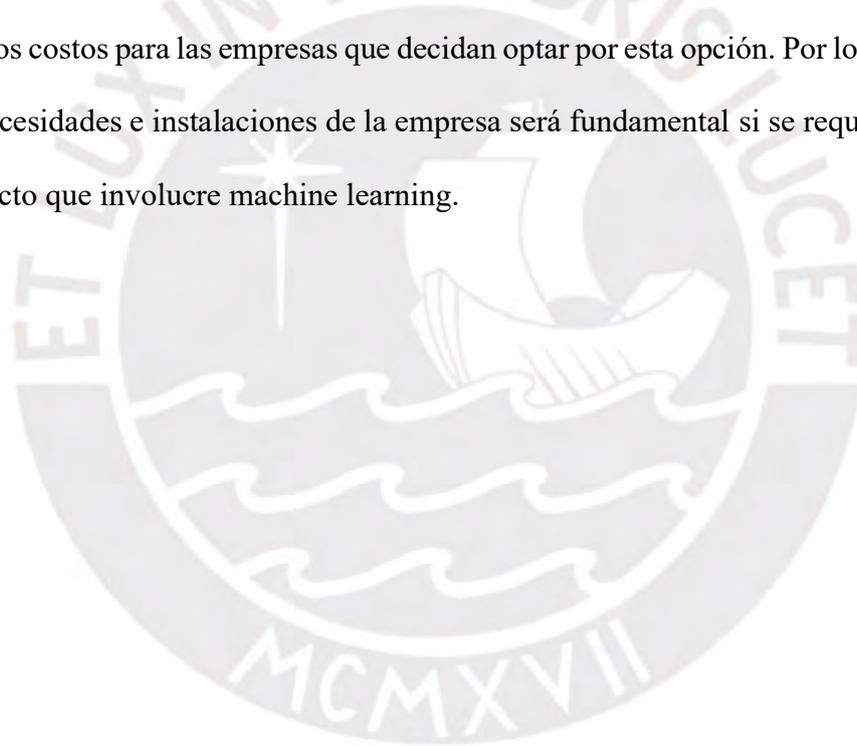
En 7 y 8 se presentan soluciones que utilizan el ensamble de varios modelos para robustecer las estimaciones; sin embargo, ambos enfoques son distintos. En 7 el ensamble se realiza asignando pesos a las predicciones de cada ensamble, y estos pesos son proporcionales a la calidad de predicciones de cada uno (evaluadas en un periodo anterior). Asimismo, también se utiliza el ensamble utilizando las estimaciones de uno de los algoritmos como variable predictora para otro más robusto. En 8, en cambio, la opción que consigue los mejores resultados se basa en la descomposición de la demanda en tendencia, estacionalidad y residuos. Para cada componente, se le aplica un modelo distinto y se combinan para obtener la predicción final. En este caso, debido a que cada modelo corresponde a un elemento de la demanda, se obtiene mayor interpretabilidad de los resultados, en comparación con la propuesta de 7, donde el ensamble puede resultar en una caja negra. Sin embargo, sería conveniente probar ambos modelos para determinar cual se acopla mejor y evaluar si es conveniente aplicar un ensamble.

## Conclusiones

Tras revisar las publicaciones presentadas en este documento, se pueden desarrollar las siguientes conclusiones al respecto:

1. Al determinar pronósticos de demanda, los modelos basados en algoritmos de machine learning, como las redes neuronales, SVM y XGBoost, suelen presentar un menor error de predicción que los métodos tradicionales. Esto se debe a que dichos algoritmos cuentan con la capacidad de recoger patrones no lineales, que son característicos de la demanda de diversos productos. Por lo tanto, la implementación de estos en reemplazo de los pronósticos clásicos generaría predicciones más precisas.
2. En el caso del entrenamiento de redes neuronales, la implementación de algoritmos novedosos como Firefly, Bat y el enfoque genético resulta en una optimización del tiempo de aprendizaje del modelo y en la precisión final de este. Si bien al algoritmo de backpropagation es el estándar, estas mejoras ayudarán a encontrar mejores parámetros y, por ende, las estimaciones serán más exactas.
3. El performance del modelo predictivo puede aumentar al combinar algoritmos como parte de la arquitectura de este. Para esto existen diversos enfoques, como el ensamble por medio de ponderación o la descomposición de la demanda (STL). Al aplicar esta técnica, es posible aprovechar los distintos algoritmos y conseguir resultados superiores al desempeño individual de estos.
4. El tratamiento de datos es una técnica con la cual los modelos predictivos descritos pueden obtener mejores resultados. La implementación de arquitecturas que incluyan la creación de nuevas variables, la reducción de dimensionalidad y el Feature Engineering demuestran la efectividad del pre procesamiento de información.

5. A pesar de los avances en este campo de estudio, el aporte de países latinoamericanos ha sido reducido hasta la fecha, pues se cuenta con pocas publicaciones académicas al respecto en estos países. Esto no implica que sea poco factible aplicar los métodos expuestos, pero se podrían obtener resultados más beneficiosos si es que se investigan las necesidades específicas en esta zona.
6. Si bien los modelos presentados suelen obtener resultados más precisos que los métodos tradicionales, esto está condicionado a contar con la suficiente información que permita a los algoritmos capturar patrones importantes. Asimismo, debido a los volúmenes de procesamiento puede ser necesario implementar arquitecturas especiales que resulten en altos costos para las empresas que decidan optar por esta opción. Por lo tanto, evaluar las necesidades e instalaciones de la empresa será fundamental si se requiere iniciar un proyecto que involucre machine learning.



## Bibliografía

- Brattin, R., Sexton, R. S., Yin, W., Wheatley, B. (2019). A neural network solution for forecasting labor demand of drop-in peer tutoring centers with long planning horizons. *Education and Information Technologies*, 24(6), 3501-3522. doi: <https://doi.org/10.1007/s10639-019-09967-3>
- Chase, R., Jacobs, F. (2018) *Operations and Supply Chain Management* (15. Ed.). New York, Estados Unidos: McGraw-Hill Education.
- Chen, Y., Wu, Q., Shao, L. (2020). Urban cold-chain logistics demand predicting model based on improved neural network model. *International Journal of Metrology and Quality Engineering*, 11(5). doi: <https://doi.org/10.1051/ijmqe/2020003>
- Goodfellow, I., Bengio, Y., Courville, A. (2016) *Deep learning*. Cambridge, Estados Unidos: MIT Press. Recuperado de <https://www.deeplearningbook.org/>
- Gupta, S., Keen, M., Shah, A., Verdier, G. (2017) *Digital Revolutions in Public Finance*. Washington, Estados Unidos: International Monetary Fund. doi: <http://dx.doi.org/10.5089/9781484315224.071>
- Gurnani, M., Korke, Y., Shah, P., Udmale, S., Sambhe, V., Bhirud, S. (2017). Forecasting of sales by using fusion of machine learning techniques. En *2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)* (pp. 93-101). IEEE. doi: 10.1109/ICDMAI.2017.8073492
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8), 2554-2558. doi: <https://doi.org/10.1073/pnas.79.8.2554>
- Ishikawa, K. (1976). *Guide to Quality Control*. Hong Kong: Asian Productivity Organization.

- Krajewski, L., Malhorta, M., Ritzman, L. (2018) *Operations Management* (11. Ed.). Pearson Education.
- Kilimci, Z. H., Akyuz, A. O., Uysal, M., Akyokus, S., Uysal, M. O., Atak Bulbul, B., Ekmis, M. A. (2019). An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain. *Complexity*. doi: <https://doi.org/10.1155/2019/9067367>
- Mostafaeipour, A., Goli, A., Qolipour, M. (2018). Prediction of air travel demand using a hybrid artificial neural network (ANN) with Bat and Firefly algorithms: a case study. *The Journal of Supercomputing*, 74(10), 5461-5484. doi: <https://doi.org/10.1007/s11227-018-2452-0>
- Ni, D., Xiao, Z., Lim, M. (2019). A systematic review of the research trends of machine learning in supply chain management. *International Journal of Machine Learning and Cybernetics*, 1-20. doi: <https://doi.org/10.1007/s13042-019-01050-0>
- Nielsen, M. (2015). *Neural networks and deep learning* (Vol. 2015). San Francisco, Estados Unidos: Determination press.
- Pawar, N., Tiple, B. (2019). Analysis on Machine Learning Algorithms and Neural Networks for Demand Forecasting of Anti-Aircraft Missile Spare Parts. En *2019 International Conference on Communication and Electronics Systems (ICCES)* (pp. 854-859). IEEE. doi: 10.1109/ICCES45898.2019.9002411
- Schwab, K. (2016) *The Fourth Industrial Revolution: what it means, how to respond*. Recuperado de <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>
- Shahrabi, J., Mousavi, S. S., Heydar, M. (2009). Supply chain demand forecasting: A comparison of machine learning techniques and traditional methods. *Journal of Applied Sciences*, 9(3), 521-527. doi: 10.3923/jas.2009.521.527

- Sharma, S., Chakraborti, S., Jha, T. (2019). Analysis of book sales prediction at Amazon marketplace in India: a machine learning approach. *Information Systems and e-Business Management*, 17(2-4), 261-284. doi: <https://doi.org/10.1007/s10257-019-00438-3>
- Skilton, M., Hovsepian, F. (2018) *The 4th Industrial Revolution*. Cham, Suiza: Springer International Publishing AG.
- Wen, R., Yan, W. (2019). Supply–Demand Prediction for Agile Manufacturing with Deep Neural Network. *Smart and Sustainable Manufacturing Systems*, 3(2), 95-105. doi: <https://doi.org/10.1520/SSMS20190025>

