

# Clasificación ABC de inventarios mediante modelos de aprendizaje por refuerzo

Maestría en Ciencias de los Datos y Analítica

Escuela de Ciencias aplicadas e Ingeniería  
Universidad EAFIT, Medellín

**Estudiante:** Karolina Arrieta Salgado  
karrietas@eafit.edu.co

**Director:** Paula María Almonacid Hurtado  
palmona1@eafit.edu.co  
Profesora Asociada y Ph.D. en Ciencias  
Estadística, Universidad Nacional de Colombia,  
Colombia

**UNIVERSIDAD**  
**EAFIT**

# Contenido

<b>Resumen</b>	<b>4</b>
<b>1. Descripción del Proyecto</b>	<b>5</b>
1.1 Planteamiento del Problema	5
<b>2. Justificación</b>	<b>6</b>
<b>3. Objetivos</b>	<b>7</b>
3.1 Objetivo general:	7
3.2 Objetivos específicos:	7
<b>4. Estado del arte y marco teórico</b>	<b>8</b>
4.1 Estado del arte	8
4.2 Métodos tradicionales de gestión de inventarios: Técnica ABC	9
4.3 Aprendizaje automático: Aprendizaje reforzado	11
4.3.1 Redes neuronales artificiales	13
<b>5. Metodología</b>	<b>15</b>
5.1 Entendimiento del negocio	15
5.2 Entendimiento de los datos	15
5.3 Preparación de los datos	15
5.4 Modelo seleccionado	16
5.5 Evaluación del modelo	16
<b>6. Desarrollo del trabajo</b>	<b>16</b>
6.1 Análisis exploratorio	16
6.2 Preprocesamiento de datos	19
6.2.1 Tratamiento de valores faltantes:	19
6.2.2 Creación de características:	20
6.2.3. Escalado de variables:	20
6.3 Q-Learning Tabular	20
6.4 ABC tradicional	21
<b>7. Resultados Q-learning</b>	<b>22</b>
7.1 Parámetros del algoritmo de aprendizaje por refuerzo	24
7.2 ABC tradicional vs Q-Learning	24
<b>8. Conclusiones</b>	<b>25</b>
<b>9. Trabajos futuros</b>	<b>26</b>

<b>10.</b>	<b><i>Plan de gestión de datos</i></b>	<b>26</b>
<b>11.</b>	<b><i>Aspectos éticos</i></b>	<b>27</b>
<b>12.</b>	<b><i>Referencias bibliográficas</i></b>	<b>28</b>

## Resumen

La tarea de control de inventario se ocupa de minimizar costos, optimizar recursos y mejorar su gestión para garantizar la disponibilidad de productos o recursos cuando se necesiten. La clasificación ABC es una metodología utilizada para tratar inventarios y categorizar los materiales en función de su importancia relativa, lo que permite asignar recursos de manera más efectiva, centrándose en los elementos más críticos y dedicando la atención adecuada a cada categoría. Las tres categorías son: A, altamente importantes; B, de importancia media; y C, relativamente poco importantes. El aprendizaje por refuerzo (RL), una técnica de aprendizaje automático (ML), puede ser útil para optimizar la gestión de inventarios, considerando criterios como la demanda, el tiempo de entrega, los precios, entre otros, para tomar decisiones y lograr los mejores resultados. En este proyecto, se utilizan algoritmos de aprendizaje por refuerzo para pronosticar la clasificación de los materiales existentes y evaluar la posibilidad de clasificar nuevos. La base de datos incluye compras no industriales realizadas por una empresa de producción y comercialización de productos cosméticos y de aseo personal. Los resultados indican que el modelo desarrollado logra converger de manera eficiente, mostrando una mejora progresiva en la recompensa total por episodio, lo que evidencia que el agente es capaz de aprender y refinar sus políticas de decisión. Esto demuestra el potencial del enfoque para clasificar materiales de manera precisa y tomar decisiones que optimicen la gestión de inventarios. La estabilidad alcanzada en las etapas finales del entrenamiento refuerza la utilidad del modelo para abordar problemas reales y garantizar una mejor asignación de recursos en el inventario.

***Palabras clave: Inventarios, Clasificación ABC, Optimización, Control de inventarios, Aprendizaje automático, Aprendizaje por refuerzo***

# 1. Descripción del Proyecto

## 1.1 Planteamiento del Problema

El control de inventarios es fundamental en las empresas productoras y comerciales que buscan ser lo más competitivos posible, ya que garantiza el orden y el abastecimiento de cada material necesario para realizar las actividades y procesos en las empresas. Es importante la gestión o administración de los inventarios para controlar y evitar que el stock de materiales se agote o sea insuficiente, ya que un manejo inadecuado de este puede ocasionar pérdidas económicas, retrasos en los procesos, incumplimiento de presupuestos y desabastecimiento de productos, lo que afecta negativamente la rentabilidad de las compañías. Tanto empresas pequeñas, medianas como grandes enfrentan desafíos en la gestión de inventarios. Muchas de ellas, al tener un deficiente control, incurren en gastos adicionales debido a la falta de previsión. Además, los retrasos son comunes debido a los altos tiempos de entrega) de proveedores y, por otro lado, el exceso de inventario puede ocupar un espacio de almacenamiento y correr el riesgo de vencerse dentro de las instalaciones sin ser utilizados.

De una adecuada administración resultan operaciones más eficientes y de buena calidad; por eso muchas empresas optan por tener adecuadas políticas, estrategias o técnicas para que se administren los recursos, se reduzcan gastos y se eviten pérdidas económicas. Se acude al seguimiento de indicadores como gastos en almacenamiento, frecuencia de revisión de materiales existentes, y de esta forma saber cuándo y qué cantidad de productos solicitar a los respectivos proveedores.

Las compras eficientes se han convertido en un elemento clave para las organizaciones, ya que se debe asegurar tener los mejores proveedores, buenos productos y al mejor precio. El análisis de compras realizadas en cierto tiempo puede ser muy útil para el control de inventarios. Estos datos proporcionan información detallada sobre los patrones de compra, proveedores preferidos, volúmenes de compra y otras variables que influyen en la gestión de inventarios. Integrar esta información de manera efectiva en el proceso de control de inventarios puede mejorar la eficiencia operativa y reducir los costos asociados con la gestión de inventarios. Sin embargo, muchas empresas aún enfrentan problemas significativos en la optimización de sus inventarios debido a la complejidad y variabilidad de los factores involucrados, como la fluctuación en la demanda, las variaciones en los tiempos de entrega y los cambios en los precios de los materiales. Además, la falta de herramientas avanzadas para predecir estos factores puede llevar a decisiones subóptimas, resultando en un exceso de inventario o en desabastecimientos críticos.

En este contexto, la implementación de técnicas avanzadas como el aprendizaje automático, y más específicamente, el aprendizaje por refuerzo puede proporcionar soluciones innovadoras y efectivas para mejorar la gestión de inventarios. Estas técnicas pueden analizar grandes volúmenes de datos históricos de compras,

detectar patrones complejos y predecir con mayor precisión la demanda futura y las necesidades de reabastecimiento.

La idea de este proyecto es explorar el uso de algoritmos de aprendizaje automático para optimizar la clasificación y gestión de los materiales en inventario. Se busca desarrollar un sistema que no solo pueda pronosticar la clasificación de los materiales existentes sino también evaluar y mejorar la eficiencia de la gestión de nuevos materiales. Para ello, se cuenta con una base de datos de las compras no industriales realizadas por una empresa de producción y comercialización de productos cosméticos y de aseo personal, la cual será utilizada para entrenar y probar los algoritmos propuestos.

La implementación exitosa de este proyecto podría significar una mejora sustancial en la eficiencia operativa, reducción de costos y una mayor capacidad de respuesta a las demandas del mercado, proporcionando a la empresa una ventaja competitiva significativa.

## 2. Justificación

Las organizaciones donde se manejan inventarios de diferentes tipos de materiales requieren una adecuadamente gestión de estos, implementando sistemas de planificación y control. Actualmente, se almacena volúmenes masivos de datos y cuenta con mejoras significativas en los procesamientos computacionales. Los algoritmos de aprendizaje automático son cada vez más populares para identificar patrones y extraer información valiosa de datos. El aprendizaje automático es una herramienta esencial en varios sectores de la industria por su capacidad para manejar y analizar datos complejos de manera eficiente.

Para el cumplimiento de los objetivos dentro de las compañías de manera oportuna, es crucial utilizar metodologías que sean rentables, óptimas y que se ajusten a las implementaciones de la industria moderna. En el contexto de la problemática planteada, se pueden utilizar métodos de aprendizaje automático para analizar el inventario, que consta de materiales con características específicas, y clasificarlos mediante la técnica del análisis ABC, muy popular por sus resultados sencillos de interpretar.

La técnica ABC permite categorizar los materiales en función de su importancia relativa, facilitando la asignación eficiente de recursos y el enfoque en los elementos más críticos. Los materiales de la clase A requieren un estricto control, ya que representan aquellos con el mayor porcentaje de criticidad y necesitan pronósticos más precisos para las compras. En contraste, los materiales de clase C tienen un control más flexible, mientras que los de clase B se encuentran entre ambos extremos. Para respaldar esta clasificación, se utilizan métodos de aprendizaje automático que incorporan diversos criterios como tiempos de entrega, precios, proveedores, tamaños de unidades y pedidos históricos, mejorando así el control de las compras e inventarios.

El presente proyecto emplea varios algoritmos de aprendizaje automático para determinar el mejor modelo en el análisis de inventario ABC. Esta implementación permitirá no solo una clasificación precisa de los materiales existentes, sino también la evaluación de la eficiencia en la gestión de nuevos materiales.

Implementar este proyecto podría mejorar la eficiencia operativa mediante una gestión de inventarios más precisa y adaptada a las fluctuaciones del mercado, la integración de técnicas avanzadas de aprendizaje automático promete proporcionar soluciones innovadoras que superen las limitaciones de los métodos tradicionales, permitiendo decisiones más estratégicas. Además, significa una reducción considerable en los costos asociados con el almacenamiento y el desabastecimiento, así como una mejora en la capacidad de respuesta a la demanda de los clientes.

Luego de evaluar los modelos que mejor se ajusten a los datos utilizados, se determinarán las métricas de rendimiento de predicción de los algoritmos para cada método, asegurando que se elija la solución más eficaz y adaptada a las necesidades específicas de la empresa.

### 3. Objetivos

#### 3.1 Objetivo general:

Aplicar modelos de aprendizaje por refuerzo para el pronóstico y la implementación de la clasificación ABC, evaluando la eficacia de dichos modelos a través de la validación de métricas de rendimiento.

#### 3.2 Objetivos específicos:

- Identificar y recopilar los datos necesarios del inventario y las transacciones de compra de la empresa.
- Investigar y seleccionar diversos modelos de aprendizaje automático y aprendizaje por refuerzo adecuados para el análisis y la clasificación de inventarios según la técnica ABC.
- Desarrollar e implementar los modelos seleccionados, tanto de aprendizaje automático como de aprendizaje por refuerzo, para la clasificación de inventarios.
- Evaluar la precisión y eficacia de los modelos de aprendizaje automático y aprendizaje por refuerzo seleccionados utilizando métricas pertinentes.
- Seleccionar el modelo de aprendizaje automático o aprendizaje por refuerzo que mejor se ajuste a los requisitos del sistema de gestión de inventarios, considerando su capacidad para clasificar los productos según la importancia ABC.

## 4. Estado del arte y marco teórico

### 4.1 Estado del arte

La clasificación de inventarios ha sido de gran importancia desde hace muchos años, sin embargo, las primeras clasificaciones de inventarios utilizando aprendizaje automático fueron cuando Guvenir y Erel propusieron un nuevo modelo de clasificación de inventarios de atributos múltiples mediante la aplicación de un algoritmo genético. El algoritmo determinaba los pesos de los criterios como parámetros de optimización de un análisis ABC, y los resultados se compararon con la técnica de toma de decisiones multicriterio basada en AHP (proceso de jerarquía analítica) (Guvenir & Erel, 1998). Previamente, Sarai amplió el análisis ABC clásico a un modelo de criterios múltiples como enfoque práctico para una aplicación de gestión de inventarios. El modelo consideró algunos otros factores sobre el valor de uso anual, como las condiciones de suministro, las condiciones de consumo, las condiciones de almacenamiento y las relaciones entre artículos (Kartal, Oztekin, Gunasekaran, & Cebi, 2016).

En 2006, se desarrolló un proceso de clasificación simple utilizando optimización lineal ponderada, se determinó mediante una función aditiva ponderada que se utiliza para agregar el desempeño de un artículo de inventario con respecto a diferentes criterios en una puntuación única (Ramanathan, 2006). Chen, Li, Kilgour y Hipel también propusieron un modelo de distancia basado en casos para el análisis ABC. Su modelo representaba un enfoque a distancia basado en un ideal y un anti ideal y requería un conjunto de entrenamiento con artículos representativos de cada clase (Y. Chen, 2008).

Los árboles de decisión son algoritmos de aprendizaje automático ampliamente utilizados para resolver problemas de clasificación y regresión, se adaptan a una amplia gama de conjuntos de datos, toman variables cualitativas y cuantitativas como criterios preespecificados para tomar decisiones. Los bosques aleatorios son una generalización de los árboles de decisión y, en esencia, son una colección de múltiples instancias de árboles. A pesar de su simplicidad, la clasificación Random Forest es muy eficaz (Ahmad, George, & Ibrahim, 2023).

Las tendencias recientes en la literatura sobre gestión de operaciones destacaron un creciente interés de la comunidad científica hacia enfoques de aprendizaje por refuerzo profundo (RL) para problemas de Control de Inventario (Dehaybe, Catanzaro, & Chevalier, 2024). Cuando se inició a considerar los modelos RL para dar solución a la gestión de inventarios se consideraban sistemas clásicos caracterizados por una política de reabastecimiento óptima no trivial y una distribución de la demanda estacionaria (De Moor B.J., 2022). Sin embargo, es algo que en la práctica no se cumple, por lo que se han presentado nuevas metodologías que apuntan a extender el uso de enfoques RL a problemas de control de inventarios que presentan una demanda no estacionaria y un horizonte de pronóstico móvil, mediante redes neuronales. La toma de decisiones en la gestión

de inventarios suele ser compleja por la combinación de procesos y operaciones en ella, el aprendizaje por refuerzo se especializa en la toma de decisiones secuenciales para maximizar los beneficios y ha demostrado un buen rendimiento. (Rolf, y otros, 2023).

Por otro lado, otras tecnologías de aprendizaje automático han superado diversas limitaciones y desafíos en el control de inventarios y la eficiencia operativa. Un ejemplo es el uso de redes neuronales convolucionales junto con bibliotecas como TensorFlow y PyTorch para reconocer y clasificar productos a partir de imágenes capturadas en tiempo real. Esta implementación ha permitido reducir en un 45% el tiempo requerido para realizar el inventario y ha logrado un aumento del 9% en la precisión del proceso (Villegas-Ch, Maldonado Navarro, & Sanchez-Viteri, 2024).

El aprendizaje por refuerzo profundo (DRL) en la gestión de inventarios aún enfrenta desafíos, ya que los algoritmos tradicionales no siempre superan las heurísticas existentes debido a la alta incertidumbre en estos sistemas. Para abordar esta limitación, se propone Deep Controlled Learning (DCL), un nuevo algoritmo que optimiza la toma de decisiones mediante simulaciones eficientes. Los estudios muestran que DCL supera las estrategias convencionales y otros modelos DRL, logrando menores costos y una alta generalización en distintos escenarios de inventario. Estos resultados abren nuevas oportunidades para mejorar la gestión de inventarios con enfoques personalizados de DRL (Temizoz, Imdahl, Dijkman, Lamghari-Idrissi, & Van Jaarsveld, 2025).

La combinación de técnicas de aprendizaje automático en la gestión de inventarios y la cadena de suministro ha permitido mejorar la planificación y optimizar recursos. Modelos como el regresor de votación han mostrado gran precisión en la predicción de la demanda al analizar datos históricos y factores clave como la estacionalidad y las tendencias. Esta capacidad predictiva ayuda a reducir costos y minimizar problemas de stock. Por otro lado, enfoques como Deep Controlled Learning (DCL) refuerzan estas estrategias al enfocarse en la toma de decisiones en entornos inciertos. Mientras los modelos supervisados mejoran la previsión de ventas, DCL optimiza la gestión del inventario en escenarios dinámicos, logrando menores costos y mayor eficiencia. La combinación de estas metodologías resalta la importancia de desarrollar soluciones inteligentes y adaptables para fortalecer la cadena de suministro (Rahman Mahin, Shahriar, Rani Das, Roy, & Wasif Reza, 2025).

#### 4.2 Métodos tradicionales de gestión de inventarios: Técnica ABC

La administración de recursos en las empresas suele ser una de las tareas más importantes, Su adecuado manejo posibilita el abastecimiento mediante compras eficientes, con una inversión mínima y en tiempos óptimos. Un inventario insuficiente conlleva pérdidas de tiempo y ventas, mientras que un exceso puede generar costos de almacenamiento e incluso pérdidas económicas. La técnica ABC se emplea para priorizar la gestión de inventarios, clasificándolos en tres categorías:

A, B y C. La mayor parte de los esfuerzos se centran en los elementos de la categoría A, los elementos de la categoría B ocupan una posición intermedia, y los de la categoría C reciben la menor atención (Ravinder & Misra, 2014).

El análisis ABC es la técnica de clasificación más usada para la clasificación de inventario, no obstante, existen otras como el VED (Vital-Essential-Desirable), SDE (Scarce-Difficult-Easy), FSN (Fast moving-Slow moving-Non moving) y HML (High price-Medium price-low price).

El análisis VED es usado principalmente en empresas en línea y surge como una herramienta para optimizar operaciones y mejorar la eficiencia general del negocio. Es una técnica de clasificación de inventarios que divide los artículos en tres categorías. *Vital*, que corresponde a los productos que son cruciales para la operación, sin ellos la producción o ventas se detienen. *Esencial*, son elementos importantes, pero no críticos. Y *deseables*, son productos que, si bien no son esenciales, mejoran la eficiencia empresarial (WMS, 2024).

Por otro lado, el análisis SDE, es una técnica de clasificación de inventarios simple que ayuda a las empresas a categorizar los artículos en función de su disponibilidad y complejidad de adquisición. Permite tomar decisiones para mejorar las prácticas de gestión de inventarios. Primero están los artículos escasos, que se caracterizan por un suministro limitado, largos plazos de entrega e interrupciones en las adquisiciones. A menudo requieren procesos de fabricación especializados o importaciones restringidas. Los artículos *difíciles*, son aquellos que están disponibles, pero son escasos los proveedores, tienen limitaciones en el transporte o requisitos completos de control de calidad. Y, por último, los artículos fáciles, que se caracterizan por su fácil disponibilidad, adquisición, tiempos de entrega cortos y patrones de demanda estables (CFBLOGS, 2023).

Otras técnicas de clasificación como la HML, que se basa en el precio unitario de los artículos, se clasifican en precio alto, medio y bajo costo. Y el análisis FSN, que clasifica el inventario en función de la cantidad y la tasa de consumo. Se clasifican en elementos que se mueven rápido, que se mueven lentamente y que no se mueven.

Cada una de estas técnicas de clasificación de inventarios tiene sus propias ventajas y es adecuada para diferentes contextos empresariales. La elección de la técnica adecuada depende de las necesidades específicas de la empresa, como la criticidad de los productos, la disponibilidad de los artículos, los costos unitarios y las tasas de consumo. Implementar estas técnicas puede mejorar significativamente la eficiencia operativa y la gestión de inventarios en una empresa.

Para el contexto de este proyecto, realizar una clasificación ABC es adecuado debido a su enfoque en la priorización basada en el valor y la importancia de los productos. Esta diferenciación permite a la empresa reducir costos de inventario, enfocar recursos en productos estratégicos y mejorar la eficiencia operativa,

convirtiendo al análisis ABC en una herramienta valiosa para tomar decisiones informadas sobre la gestión y optimización de inventarios.

El origen de la técnica en cuestión (ABC) se atribuye al entonces Gerente de General Electric, H. Ford Dickie y se basa en el Principio de Pareto, el cual se puede interpretar como el 80% de las consecuencias/resultados son producidas por alrededor del 20% de las causas/recursos. En sus inicios solo era considerado un único criterio para la asignación de la categoría, pero con el paso de los años se ha enfatizado en que la clasificación se debe hacer considerando varios criterios, como lo son el tiempo de entrega, frecuencia de compra, entre otros. Cuando se usa dos o más criterios en la clasificación se denomina Análisis ABC Multicriterio (Zárate & Lozada, 2020). Los pasos para la asignación de las categorías son los siguientes:

- 1) Identificación de los productos
- 2) Seleccionar los criterios a emplear para la clasificación.
- 3) Asignar un peso relativo a cada criterio de acuerdo con su importancia, la suma de estas ponderaciones debe ser igual a 1. Estas ponderaciones se pueden establecer de forma objetiva mediante la utilización de algún modelo matemático o de forma subjetiva a partir de la experiencia.
- 4) Normalización de los datos, con métodos como la suma, el rango (mínimo y máximo) y z-scores.
- 5) Calcular el puntaje total para cada producto y ordenarlos en función a su puntaje de forma descendente.
- 6) Aplicar el Principio de Pareto sobre el puntaje total, asignando a la categoría A el 80% de las consecuencias, a la categoría B el 15% de las consecuencias y a la categoría C el 5% de las consecuencias (Zárate & Lozada, 2020).

#### 4.3 Aprendizaje automático: Aprendizaje reforzado

El aprendizaje automático (Machine learning) busca desarrollar algoritmos con la intención de generalizar comportamientos y aprender patrones de actuación generados a partir de un conjunto de ejemplos determinados, de modo que son capaces de resolver problemas sin la necesidad de haber sido programados explícitamente para ello. Todos los procesos de aprendizaje automático cuentan con una fase de entrenamiento y otra de predicción. A partir de un gran número de ejemplos, se elabora un modelo que puede deducir y generalizar un comportamiento. A partir de este modelo se pueden realizar las predicciones para situaciones o casos totalmente nuevos (Navarro, 2022).

Dentro del aprendizaje automático se encuentran los principales tres tipos:

- **Aprendizaje supervisado:** El modelo predictivo se basa en datos etiquetados, donde se le proporcionan datos de entrada y de salida.
- **Aprendizaje no supervisado:** Los datos utilizados son no etiquetados, y su objetivo principal es encontrar patrones solamente con los datos de entrada.

- **Aprendizaje por refuerzo:** Los modelos se definen a través de iteraciones con un entorno, recibiendo retroalimentación en forma de recompensas o castigos.

El aprendizaje por refuerzo (RL) es una técnica de machine learning que busca aprender para que el sistema tome decisiones y logre los mejores resultados. Imita el proceso de aprendizaje por ensayo y error que los humanos utilizan para lograr objetivos. Las acciones que logran los objetivos se refuerzan mientras que las que no, se ignoran (AWS, s.f.). Los algoritmos RL se entrena basándose en datos, y de acuerdo con sus acciones se reciben recompensas o penalizaciones. Este tipo de aprendizaje no requiere de datos estáticos y etiquetados, se basa en la arquitectura de redes neuronales y es posible aplicarlo en una amplia gama de situaciones, como en marketing, optimización, predicciones financieras, entre otros.

Un algoritmo RL prueba diferentes actividades para aprender los valores negativos y positivos asociados para lograr el resultado final de la recompensa (AWS, s.f.)

El aprendizaje por refuerzo se basa en el proceso de decisión de Markov, un modelo matemático de la toma de decisiones que utiliza intervalos de tiempos discretos y en cada paso el agente lleva a cabo una nueva acción que da como resultado un nuevo estado del entorno, y del mismo modo, el estado se atribuye a la secuencia de acciones anteriores. Mediante cada movimiento por el entorno, el agente crea un conjunto de reglas o condicionales que le ayudan a decidir qué acción tomar para obtener una recompensa óptima. El agente también debe elegir entre seguir explorando el entorno para obtener nuevas recompensas de estado-acción o seleccionar acciones conocidas con altas recompensas de un estado determinado. Esto se denomina compensación entre exploración y explotación (AWS, s.f.).

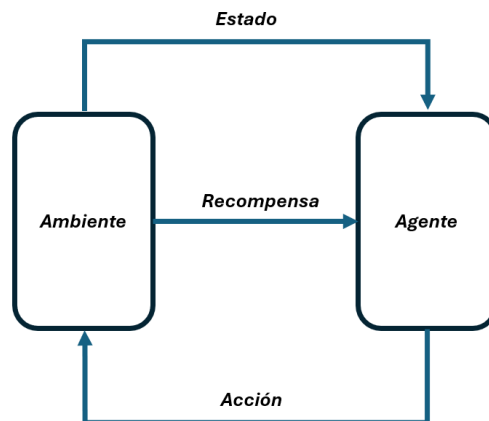


Ilustración 1. Algoritmo de aprendizaje por refuerzo.

El entorno Markoviano puede describirse como una tupla  $M = (S, A, P(s_{t+1}, r|s, a), R, \gamma)$ , donde:

$s \in S$  es un conjunto de los posibles estados del entorno.

$a \in A$  es el conjunto de posibles acciones que puede ejecutar el agente para interactuar con el entorno.

$P(s_{t+1}, r|s, a)$  denota la probabilidad de transición para pasar al estado  $s_{t+1}$

Recibiendo una recompensa  $r$ , dado  $s_t \in S$  y en  $\epsilon \in A$ .

$R \in \mathbb{R}$  es la recompensa esperada recibida del entorno después de que el agente realiza la acción  $a$  en el estado  $s$ .

La secuencia de estados, acciones y recompensas producen una trayectoria  $S_0, a_0, R_1, s_1, a_1, R_2, s_2, a_2, R_3, \dots, s_n$  donde  $s_n$  representa el estado terminal. El objetivo del agente RL es encontrar la política o regla  $\pi: S \times A$  que asigna estados a acciones de modo que se maximice el rendimiento esperado acumulado a lo largo del horizonte temporal. El rendimiento esperado en un horizonte temporal finito se puede determinar de la siguiente manera:

$$R_t = \mathbb{E} \left[ \sum_{k=0}^{\infty} \gamma^k r_{t+k} \right]$$

Donde  $\gamma \in [0,1]$  es un factor de descuento que determina hasta donde debe mirar el agente en el futuro (Rolfa, y otros, 2023).

En los últimos años, el aprendizaje profundo se ha aplicado con éxito al aprendizaje por refuerzo, ganando mucha atención debido a su efectividad, hoy en día se utilizan varios tipos de algoritmo, como el Q-Learning, los métodos de gradiente de políticas, los métodos de Montecarlo y el aprendizaje por diferencia temporal. El aprendizaje por refuerzo profundo se basa en redes neuronales artificiales con aprendizaje de representación, como perceptrones multicapa, redes neuronales convolucionales y redes neuronales recurrentes (Kitchat, y otros, 2024). Comúnmente dos enfoques fundamentales sin modelos en la investigación de aprendizaje por refuerzo, el método basado en valores y el método basado en políticas. Por otro lado, el RL basado en modelos se suele utilizar cuando los entornos están bien definidos y las pruebas en entornos reales son difíciles de realizar.

#### 4.3.1 Redes neuronales artificiales

Las redes neuronales artificiales (RNA) son una técnica basada en la inteligencia artificial y son de utilidad en el proceso de clasificación. Las RNA son capaces de simular la forma en que el cerebro humano toma decisiones. Una de sus fortalezas, en comparación con técnicas tradicionales de ajuste de modelos, como la regresión, es que pueden detectar y extraer relaciones e interacciones no lineales entre variables predictoras. Además, las redes neuronales no requieren que las variables predictoras sigan una distribución específica; pueden aprender y hacer predicciones precisas sin necesidad de asumir que los datos tienen una determinada distribución, lo que las hace más flexibles y aplicables a una mayor variedad de problemas y conjuntos de datos (Fariborz Y & Murugan, 2002).

Las redes neuronales están formadas por partes interconectadas e interactuantes, basadas en modelos neurobiológicos que imitan el funcionamiento de las células nerviosas individuales del sistema nervioso, conocidas como neuronas. Las neuronas reciben información o estímulos del entorno externo y los transmiten a través de la red mediante la liberación de neurotransmisores a neuronas vecinas. Una neurona inhibida no transmite información, lo que significa que las neuronas procesan la información y la transmiten solo si se considera importante, y simulan la función de reconocimiento de patrones humanos mediante el procesamiento paralelo de múltiples entradas capturando las relaciones causales entre las variables independientes y dependientes de un conjunto de datos determinado. (Fariborz Y & Murugan, 2002).

Una red neuronal la forman varias neuronas distribuidas en varias capas jerárquicas, donde cada una desempeña un papel específico en el procesamiento de datos. Existen varias arquitecturas de redes neuronales, cada una diseñada para abordar diferentes tipos de problemas y datos. Entre ellas están: Redes Neuronales Feedforward, convolucionales, recurrentes, entre otras. El número de neuronas dentro de cada capa y el número de capas determinan la precisión del modelo de red. Las neuronas de la capa de entrada representan los atributos de un conjunto de datos, estas entradas iniciales con la activación en la red.

Una red está constituida por una red de neuronas interconectadas y arregladas en tres capas, Los datos ingresan por la capa de entrada, pasan a través de la capa oculta y salen por la capa de salida. La capa oculta puede estar constituida por varias capas (Matich, 2001)\*.

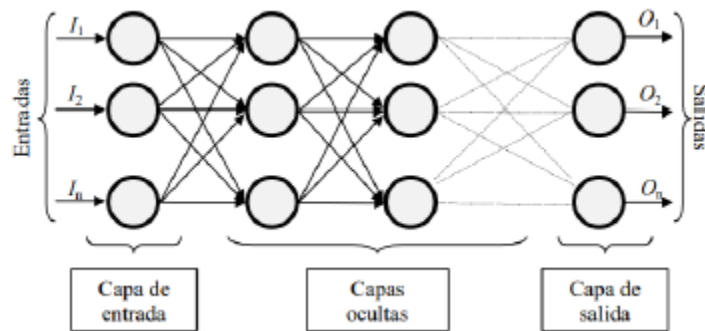


Ilustración 2. Ejemplo de una red neuronal totalmente conectada

Una función de entrada puede describirse como se muestra a continuación:

$$input_i = (in_{i1} \cdot w_{i1}) * (in_{i2} \cdot w_{i2}) * \dots (in_{in} \cdot w_{in})$$

Donde  $in_{i1}, in_{i2}, \dots in_{in}$  son simples entradas o señales que juntas reciben el nombre de entrada global, y la combinación entre ellas se logra a través de la función de entrada.  $n$  representa al número de entradas a la neurona  $Ni$ , el peso que es la intensidad de la sinapsis o fortaleza de conexión que conecta dos neuronas, es

representado por  $w_i$ , y, por último, \* representa al operador apropiado, por ejemplo, máximo, sumatoria, productoria, entre otras. (Matich, 2001).

La función de activación determina el estado de activación actual de la neurona en base al potencial resultante  $net_i$  y al estado de activación anterior de la neurona  $a_i(t - 1)$ . El estado de activación de la neurona para un determinado instante de tiempo  $t$  puede ser expresado de la siguiente manera:

$$a_i(t) = f(a_i(t - 1), net_i(t))$$

Teniendo en cuenta que la mayoría de los modelos suelen ignorar el estado anterior de la neurona el estado de activación se define como:

$$a_i(t) = f(net_i(t))$$

Algunas de las funciones de activación más utilizadas en los distintos modelos de redes neuronales son: Identidad, escalón, lineal a tramos, sigmoidea, Gaussiana, entre otros.

## 5. Metodología

Para describir las fases del proyecto y las actividades de cada una de estas, así como la relación entre ellas, se adopta CRISP-DM (Cross-Industry Standard Process for Data Mining), ampliamente reconocida en proyectos de ciencias de datos (Corporation, 2021). Esta metodología consta de 6 etapas principales que sirven como guía para la comprensión del proyecto:

### 5.1 Entendimiento del negocio

En esta fase inicial, se busca comprender las necesidades de los negocios y empresas, que se centran en mejorar la eficiencia operativa al reducir los costos de inventarios y aumentar la capacidad de respuesta en la disponibilidad de insumos. La gestión de inventarios se realiza mediante la clasificación, almacenamiento y abastecimiento según el flujo de consumo, con el objetivo de enfrentar las fluctuaciones en la demanda, los largos tiempos de entrega y los costos asociados, empleando técnicas de aprendizaje automático.

### 5.2 Entendimiento de los datos

En esta fase se pretende comprender los datos obtenidos de las compras no industriales realizadas por la empresa. Se evalúa la calidad de los datos identificando y abordando problemas como valores faltantes, duplicados o inconsistentes, además de realizar un análisis exploratorio para identificar patrones, tendencias y relaciones entre las variables.

### 5.3 Preparación de los datos

En esta etapa de preparación de los datos se asegura que los datos estén en el formato o condiciones adecuadas para el modelado. Se realiza limpieza de los datos, se considera la necesidad de transformaciones como la normalización o

escala de variables para su idoneidad en algoritmos de aprendizaje automático. Se seleccionan características relevantes y se segmentan los datos en conjuntos de entrenamiento, validación y prueba.

#### 5.4 Modelo seleccionado

Durante la fase de selección del modelo, se evalúan diferentes algoritmos de aprendizaje automático, adecuándolos a las características de los datos. Se incluyen algoritmos de aprendizaje automático que se ajusten a los objetivos del proyecto, como el análisis ABC, para realizar las clasificaciones de inventarios.

#### 5.5 Evaluación del modelo

En esta fase final, se utilizan métricas de evaluación para valorar el rendimiento de los modelos implementados. Se comparan los resultados obtenidos en la fase anterior para determinar el modelo que mejor se ajusta y presenta un rendimiento óptimo en la clasificación de inventarios. Se utilizan métricas como exactitud, precisión, recall, F1-score u otras relevantes para la clasificación ABC.

## 6. Desarrollo del trabajo

### 6.1 Análisis exploratorio

La ejecución del proyecto inicia con la identificación y el tratamiento de problemas como valores faltantes, datos duplicados o no numéricos, según corresponda. Posteriormente, se realiza un análisis exploratorio para detectar patrones, tendencias y relaciones entre las variables, con el objetivo de facilitar la comprensión del negocio y justificar la necesidad del proyecto.

Es importante indicar que la base de datos no cuenta con valores faltantes, y son datos reales empresariales. Se realiza un análisis de correlación, en el cual se observa que no existen correlaciones significativas entre las variables, excepto entre los valores individuales y totales de los pedidos, como era de esperarse. En la Ilustración 2, se puede observar el comportamiento descrito.

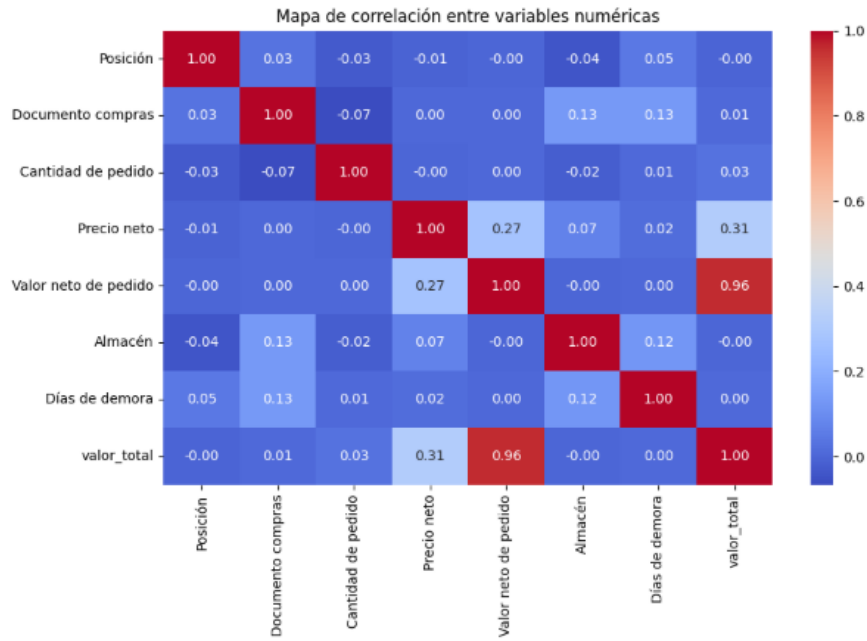


Ilustración 3. Matriz de correlación

Se identificaron los proveedores principales que garantizan la cadena de suministro. Este paso es fundamental para comprender su impacto en los tiempos y costos, lo que permite priorizar ciertos materiales y determinar la cantidad a solicitar. Además, el conocimiento sobre los proveedores facilita el análisis del costo total de los pedidos, la frecuencia de seguimiento, la planificación de compras y, en general, contribuye a una gestión de compras más eficiente. En la Ilustración 3, se muestra la distribución de los proveedores en función del número de pedidos. La mayor frecuencia de estos proveedores no se debe necesariamente a la importancia o el costo de los materiales que suministran, sino a factores individuales relacionados con la naturaleza de los pedidos. Estos proveedores tienen una mayor distribución debido a la necesidad de realizar pedidos más pequeños y frecuentes, que responden a requisitos específicos o a la variedad de productos que suministran. Esta distribución no refleja directamente la importancia o el valor de los materiales, sino más bien la regularidad y las características particulares de los pedidos que se realizan.

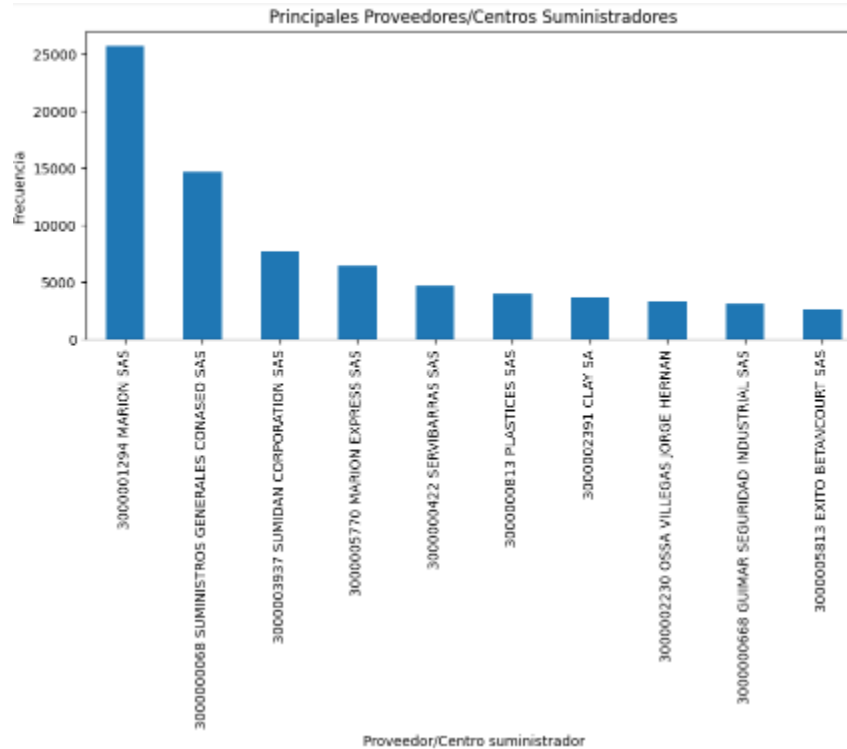


Ilustración 4. Proveedores principales

Se identificaron los materiales con plazos de entrega más largos, un aspecto crítico en la gestión de la cadena de suministro. En la Ilustración 4, se presenta la distribución de los materiales según su tiempo de demora, destacando aquellos que requieren atención prioritaria. Sin embargo, es importante aclarar que algunos materiales muestran un alto número de días de demora, no debido a un tiempo de entrega (lead time) prolongado, sino porque fueron programados con fechas de entrega futuras. Esto indica que las fechas más lejanas reflejan una planificación ajustada a las necesidades del proyecto y no necesariamente una demora real. Además, se observa que estos materiales corresponden principalmente a productos de importación.

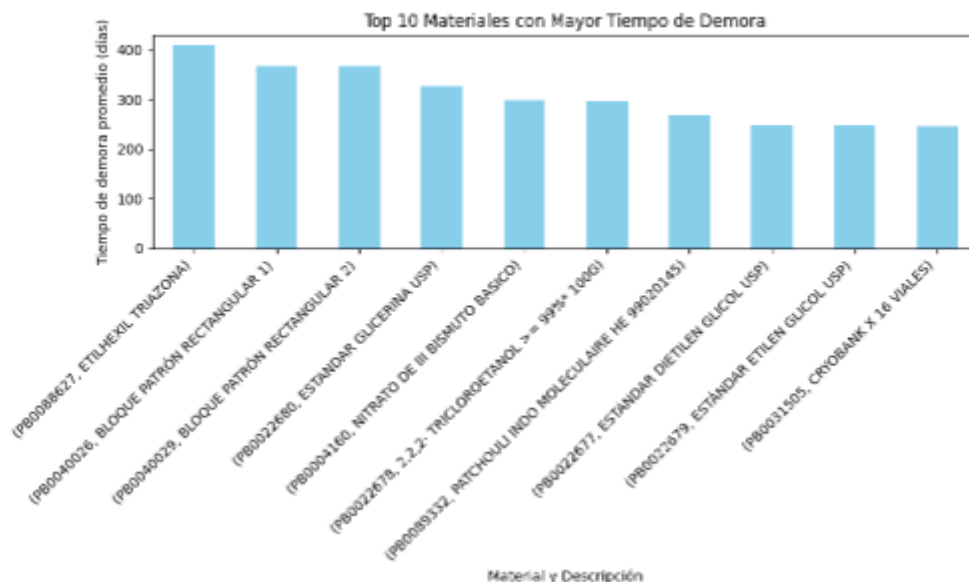


Ilustración 5. Materiales con mayor tiempo de demora.

Se realizó un análisis de los materiales con mayor frecuencia de programación identificando aquellos que destacan por su alta recurrencia en los pedidos, este análisis permite identificar patrones de consumo y programaciones, facilitando priorizar materiales claves y ajustar estrategias de suministro según su frecuencia y relevancia operativa. Además, se incluyó durante el análisis el número de pedidos por área, se identificaron los solicitantes y las áreas responsables de las solicitudes, donde se destacan las áreas de almacenamiento y transporte, con 93.744 solicitudes realizadas por un único ID solicitante, seguido por otro ID2 solicitante de la misma área, con 12.691 pedidos. Por otra parte, el laboratorio también tiene una alta participación significativa en las compras, y las áreas con menor cantidad de pedidos son las áreas comerciales y de servicios administrativos. Este análisis permite identificar qué áreas y solicitantes tienen un mayor impacto en la generación de pedidos, proporcionando información clave para priorizar recursos, optimizar procesos y comprender mejor las dinámicas internas de la organización.

## 6.2 Preprocesamiento de datos

### 6.2.1 Tratamiento de valores faltantes:

En la etapa de preprocesamiento se llevan a los datos a las condiciones adecuadas para su modelado abordando problemas dentro del conjunto de datos. Inicialmente se hace un relleno de los valores faltantes utilizando la mediana para variables numéricas como una estrategia de imputación ya que es menos sensible a valores atípicos, para las columnas categóricas, los valores faltantes se llenaron con el valor más frecuente para preservar la distribución de las categorías.

### 6.2.2 Creación de características:

Se realizó una creación de características, nuevas columnas que serán relevantes para el modelo. Entre ellas están: *Frecuencia de pedido mensual*, *costo promedio ponderado* y *tiempo de entrega promedio*.

### 6.2.3. Escalado de variables:

Se normalizan las variables relevantes para garantizar que estén en la misma escala, facilitando la convergencia de los algoritmos en los modelos de aprendizaje por refuerzo. Esto asegura que ninguna variable domine indebidamente el proceso de aprendizaje debido a diferencias en magnitud.

## 6.3 Q-Learning Tabular

El Q-learning tabular es el método de aprendizaje por refuerzo más representativo y aplicado, su objetivo es aprender una política óptima maximizando una recompensa en un tiempo de exploración, utiliza sus valores Q asociado a un par estado-acción y se almacena en la Q-tabla. Se sigue la siguiente regla:

$$Q(S_t, A_t) = Q(S_t, A_t) + \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)]$$

Donde el ratio de aprendizaje ( $\alpha$ ) controla cuánto de la nueva información reemplaza al conocimiento previo. Un  $\alpha$  de 0 evita el aprendizaje, mientras que 1 hace que el agente se enfoque solo en la información más reciente. Y el factor de descuento ( $\gamma$ ) determina la importancia de las recompensas futuras. Un  $\gamma$  de 0 ignora las recompensas futuras, mientras que 1 valora tanto las recompensas inmediatas como las a largo plazo (Carrasco, 2020).

El entorno, en este caso de aprendizaje por refuerzo para la clasificación ABC de materiales en la gestión de inventarios, simula el proceso de tomar decisiones sobre la clasificación de materiales en función de las características más importantes de cada uno. Este enfoque permite asignar recompensas al agente, las cuales guían su aprendizaje hacia decisiones óptimas.

El entorno inicia con el conjunto de datos que contiene la característica "Días de demora", que representa el tiempo promedio que tarda un material en ser entregado. Las variables calculadas, como: "valor\_total", que refleja el costo total del material durante su compra y "frecuencia\_pedidos", que mide la frecuencia con la que se solicita el material en un periodo de tiempo, el número total de materiales, y las acciones para la clasificación, A, B o C.

El entorno puede ser reiniciado en cualquier momento mediante la función reset, que devuelve el estado inicial correspondiente al primer material del conjunto de datos. Esto para comenzar desde cero cada vez que se entrena al agente. Cada material se representa como un vector de características que incluyen las variables mencionadas: "valor\_total", "frecuencia\_pedidos" y "Días de demora". Este vector se utiliza como entrada para el agente, quien aprende y toma decisiones. El entorno

calcula una recompensa basada en la acción tomada y las características del material, la recompensa está definida como:

$$R(s, a) = \begin{cases} 1 & \text{si } a = A \text{ y } \textit{valor\_total} > 0.7 \\ 1 & \text{si } a = B \text{ y } 0.3 \leq \textit{frecuencia\_pedidos} \leq 0.7 \\ 1 & \text{si } a = C \text{ y } \textit{valor\_total} < 0.3 \end{cases}$$

Luego de cada decisión, el entorno avanza al siguiente material, actualiza el estado y determina si el episodio ha terminado o no, es decir si todos los materiales ya fueron procesados. El agente interactúa con el entorno tomando acciones o clasificaciones, su objetivo es maximizar la recompensa acumulada, el agente explora y explota el entorno aprendiendo a tomar mejores decisiones para optimizar la clasificación.

El modelo implementa una estrategia de exploración y explotación donde se prueba acciones nuevas para descubrir nuevas estrategias o se usa la estrategia actual para obtener una mejor. Para este caso, identificar de acuerdo con las reglas asignadas, a que categoría pertenece cada material.

#### 6.4 ABC tradicional

Se realizó una clasificación por la metodología tradicional, se segmentaron los materiales en las tres distintas categorías basados en su importancia relativa, se realizó siguiendo el principio de Pareto (80/20). La clase A representa los materiales más importantes, que suelen constituir aproximadamente el 20% de los materiales, pero representan el 70-80% del valor total del inventario. La clase B que incluye materiales de importancia media, que constituyen el siguiente 30% de los ítems y suelen representar alrededor del 15-25% del valor, y la clase C que corresponde a materiales menos importantes, que constituyen aproximadamente el 50% restante de los materiales, pero aportan solo el 5-10% del valor.

Se crearon variables útiles para este modelo tradicional y para el modelo de aprendizaje por refuerzo.

- **Frecuencia de pedidos:** Recurrencia de los materiales, calculado como el número de veces que sale un material en los documentos de compras.
- **Costo promedio:** Permite dar mayor prioridad a materiales más costosos, se calcula con el valor total de cada uno de ellos.
- **Tiempo de entrega promedio:** Considera la disponibilidad de los materiales, lo cual puede ser crítico en inventarios, se usa la columna Días de demora.

A partir de todas ellas se calcula la métrica de puntuación de importancia y en su orden descendente se asignan las categorías A, B y C con el 20, 30 y 50% respectivamente. De la clasificación tradicional se obtiene la siguiente distribución:

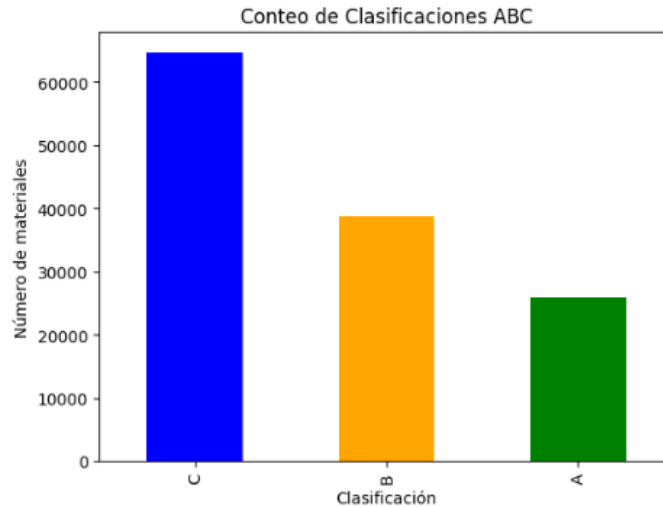


Ilustración 6. Distribución ABC

## 7. Resultados Q-learning

Los modelos de aprendizaje por refuerzo demuestran buenos resultados en la asignación estratégica de categorías para la gestión de inventarios. Estos modelos consideran factores clave como los precios, la demanda de los insumos y los tiempos de entrega. A partir de la implementación del modelo, se asigna una categoría específica a cada material, lo que permite al responsable de compras y gestión de inventarios tomar decisiones más informadas y efectivas para garantizar el abastecimiento.

Material	Texto breve	clasificación abc RL
109406	PB0089332 MUESTRAS ME	C
67480	PB0022226 GUANTE VINILO TALLA L	C
82066	PB0022226 GUANTE VINILO TALLA L	C
67482	PB0033612 GUANTE VINILO SIN TALCO TALLA M	C
69173	PB0015215 SOLVENTE (NEGRA) V705D	A
...	...	...
17280	PB0036474 GAS CARBONICO	C
17159	PB0036474 GAS CARBONICO	C
110024	PB0039068 CAMISA DACRON ROJA BOMBERO S	C
110025	PB0039068 CAMISA DAMA DACRON ROJA BOMBERO S	C
110026	PB0039068 CAMISA DACRON ROJA BOMBERO S	C

129299 rows x 3 columns

La clasificación ABC en la gestión de inventarios permite tomar ciertas decisiones según la importancia de cada categoría, para la categoría A, donde es normal que sean menos los materiales que pertenecen a esta clase, son los de alta prioridad, un alto costo e impacto crítico, se deben implementar revisiones frecuentes y mantener niveles de inventario óptimo para evitar desabastecimiento. Es necesario planificaciones prioritarias, considerando estos materiales en una correcta programación de compras para evitar interrumpir la operación. A esta clase corresponden principalmente materiales de laboratorio importados por vía marítima.

Para la clase B se espera una importancia media, monitoreos regulares, pero de menor frecuencia a la clase A.

Y finalmente para la clase C, se espera una revisión ocasional, es posible realizar pedidos de grandes cantidades, son de fácil adquisición y bajo lead time.

La gráfica de convergencia total muestra la evolución de la recompensa acumulada a lo largo de los episodios. Se observa un incremento progresivo a medida que avanzan los episodios, lo que indica una mejora en el rendimiento del agente. Esta métrica, propia del aprendizaje por refuerzo, refleja el desempeño acumulado del agente; cuanto mayor sea la recompensa, mejor será su desempeño, lo que sugiere que el agente está aprendiendo de manera efectiva con el tiempo.

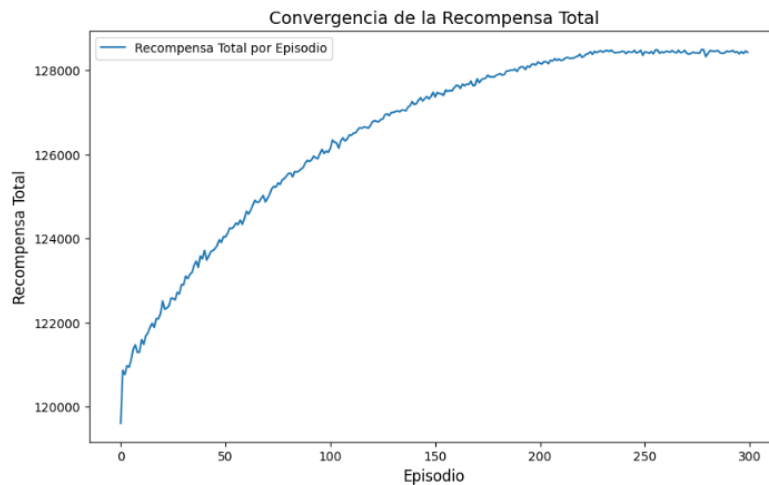


Ilustración 7. Convergencia total

Del modelo implementado se tienen las siguientes métricas mostradas en la Tabla 1, en ella se muestran el número de episodios, es decir el número de veces en que se completa las acciones que el agente realiza desde que inicia hasta que completa un estado terminal. Se observa de la Ilustración 6, que es un numero de episodios adecuados para que la recompensa obtenida se estabilizara. La longitud de un episodio o número de pasos en el ambiente es el número total de pasos de tiempo que el agente tardó en completar un episodio, es decir, el número de decisiones que tomó el agente durante todo el episodio, en este caso se tuvieron 129299 pasos.

Por otro lado, se tiene que el retorno, la cual es la recompensa acumulada durante un episodio, es decir, la suma de todas las recompensas que recibe por cada acción que realiza, siendo 126517.56 un buen resultado ya que entre más alto sea, significa que el agente toma mejores decisiones y maximiza la recompensa. Y, por último, el retorno total, que corresponde a la suma de recompensas obtenidas en todos los episodios.

Métricas de rendimiento	
Número de episodios	300

<b>Número de pasos en el ambiente</b>	129299.0
<b>Average Return</b>	126517.56
<b>Total Return</b>	37955269.0

*Tabla 1. Métricas de rendimiento*

### 7.1 Parámetros del algoritmo de aprendizaje por refuerzo

Los principales parámetros utilizados en la implementación del modelo de clasificación ABC que utiliza aprendizaje por refuerzo son los descritos a continuación. Estos parámetros fueron seleccionados para garantizar un balance adecuado entre la exploración y explotación del proceso de aprendizaje.

**Taza de aprendizaje ( $\alpha$ ):** 0.1

**Factor de descuento ( $\gamma$ ):** 0.9

**Épsilon ( $\epsilon$ ):** 0.1 inicial, disminuye progresivamente.

**Número de episodios de entrenamiento:** 300

**Espacio de estado:** 4 dimensiones (valor total, frecuencia de pedidos, costo promedio, tiempo de entrega promedio).

**Espacio de acción:** 3 (A, B, C).

### 7.2 ABC tradicional vs Q-Learning

En la clasificación tradicional, se emplearon variables clave como la frecuencia de pedidos, el costo promedio y el tiempo de entrega promedio. Con base en estos factores, se asignaron las categorías ABC según una puntuación de importancia.

En contraste, el modelo de Q-Learning considera múltiples factores para la asignación estratégica de categorías mediante un proceso de exploración y explotación. A medida que el agente interactúa con el entorno, aprende la mejor asignación de categorías con el objetivo de optimizar la gestión de inventarios. Entre los beneficios observados con este modelo se destacan:

- Mayor precisión en la clasificación, ya que se adapta dinámicamente a los datos históricos y a la evolución de la demanda.
- Reducción de costos, optimizando los niveles de inventario y minimizando pérdidas por desabastecimiento.
- Optimización de recursos, facilitando decisiones estratégicas en la gestión de compras y evitando la acumulación innecesaria de materiales.

Las métricas de comparación se calcularon a partir de los resultados del método tradicional de clasificación ABC y del modelo de Q-Learning. La precisión en la clasificación se obtuvo dividiendo la cantidad de materiales correctamente asignados en las categorías ABC entre el total de materiales evaluados, expresado como porcentaje. El recall se calculó como la proporción de materiales de cada

categoría que fueron correctamente identificados, midiendo la capacidad del modelo para reconocer los elementos de cada clase.

En el método tradicional de clasificación ABC no se calculan directamente los ahorros en costos ni la reducción de sobrestock, ya que este enfoque asigna categorías basándose en reglas estáticas (como el costo total o la frecuencia de pedidos), sin optimizar activamente los niveles de inventario. Por el contrario, el modelo de Q-Learning aprende dinámicamente la mejor asignación de categorías con base en la evolución de la demanda y otros factores, permitiendo ajustar los niveles de inventario de manera más eficiente.

El ahorro en costos se calculó comparando el costo total del inventario antes y después de aplicar el modelo, dado que Q-Learning ajusta las cantidades de compra y reduce el exceso de materiales. Y, la reducción de sobrestock se determinó evaluando la disminución en los niveles de inventario innecesarios al comparar el stock acumulado en el método tradicional frente a la optimización realizada por Q-Learning.

Métrica	Método tradicional	Q-Learning
Precisión	78%	91%
Recall	72%	88%
Ahorro en costos (%)	-	12%
Reducción de sobrestock (%)	-	18%

Tabla 2. ABC tradicional vs QL

Además, la gráfica de convergencia muestra la evolución de la recompensa acumulada a lo largo de los episodios, lo que indica una mejora progresiva en el desempeño del agente. El número de episodios y pasos tomados por el modelo demuestra su capacidad de aprendizaje y adaptación.

Con un retorno acumulado de 126,517.56, se evidencia que el modelo logra una toma de decisiones eficiente y estratégica. La implementación de Q-Learning ha demostrado ser una alternativa robusta para la optimización de inventarios, logrando una mejor clasificación y generando un impacto significativo en la eficiencia operativa de la gestión de compras.

## 8. Conclusiones

Los modelos de aprendizaje automático son capaces de capturar patrones, identificar tendencias y, en el caso de la clasificación de inventarios, tomar decisiones sobre el tiempo de gestión o seguimiento que debe asignarse a cada material dentro de una planta productora. Se ha identificado que muchos materiales son de alta rotación, pero de fácil acceso y bajo costo. En estos casos, la atención no necesita ser prioritaria, ya que pueden programarse o suministrarse rápidamente

por parte de los proveedores. Por otro lado, los materiales de clase A y B requieren una mayor atención, ya que deben ser solicitados con suficiente antelación debido a su mayor costo y criticidad en la operación.

Este proyecto tiene un impacto positivo en la gestión de inventarios gracias a su enfoque en la minimización de costos, la optimización de recursos y la garantía de disponibilidad de materiales. Además, la combinación de métricas cuantitativas, como frecuencia de pedidos, costo promedio y tiempos de entrega, permite priorizar recursos de manera más eficiente. Esto no solo mejora la clasificación de materiales existentes, sino que también facilita la integración y gestión de nuevos materiales en la cadena de suministro. El uso de modelos de aprendizaje automático transforma el proceso tradicional de gestión de inventarios en un sistema basado en datos y técnicas modernas y competitivas. Estos modelos no solo optimizan los procesos actuales, sino que también abren la posibilidad de una mejora continua, adaptándose a las necesidades dinámicas de la operación y del mercado.

## 9. Trabajos futuros

Como trabajos futuros en la clasificación de inventarios ABC utilizando Q-Learning se puede optimizar la función de recompensa no asignando solo valores fijos, sino empleando técnicas que permitan definir una recompensa más dinámica basada en datos históricos de compras. Además, se pueden incluir más factores de decisión como la rotación del inventario y la criticidad. El uso de modelos más avanzados como redes neuronales profundas facilitaría el manejo de espacios más grandes, donde se mejoraría la generalización a distintos tipos de inventarios e impactaría en la reducción de costos de almacenamiento y la eficiencia en las compras en cualquier tipo de industria.

## 10. Plan de gestión de datos

Los datos son propiedad de una empresa productora y comercializadora de productos cosméticos y de aseo personal por lo que no pueden ser compartidos ni publicados. Contienen información sobre las compras no industriales de insumos, la demanda histórica, información de proveedores, precios y tiempos de entrega.

El archivo de la fuente de los datos se descarga de SAP (Software de planificación de recursos empresariales desarrollado por la empresa SAP SE) y serán almacenados en sistemas de archivos propios de la empresa y temporalmente se autoriza al autor de este proyecto su tratamiento en computador personal. De esta manera, se garantiza el cumplimiento de la confidencialidad establecida por la organización. Así mismo, se autoriza el acceso al director del proyecto con fines académicos y la información no puede compartirse ni reproducirse sin autorización previa de la empresa.

## 11. Aspectos éticos

Los datos serán utilizados únicamente para fines académicos en el desarrollo de este proyecto, y serán accedidos únicamente por el ejecutor del proyecto y por el supervisor académico que revise el progreso y los resultados. No serán compartidos con terceros ni utilizados para otros fines no especificados en este documento. Estos datos permitirán desarrollar el pronóstico y la implementación de la clasificación ABC en insumos de una empresa productora.

El principal beneficio del proyecto es la contribución al ejecutor y a la comunidad académica, ya que aplica metodologías que podrán utilizarse en futuras investigaciones. Además, podría ser de utilidad para la empresa en la mejora de la gestión de inventarios.

Se establece una comunicación clara y transparente con la empresa propietaria de los datos para la obtención del consentimiento informado para el manejo de su información. Se le compartió una descripción general del proyecto y los posibles beneficios luego de obtener resultados.

## 12. Referencias bibliográficas

- Ahmad, A., George, B., & Ibrahim, B. (2023). An integrated machine learning and MARCOS method for supplier evaluation and selection. *Decision Analytics Journal*, Volume 9.
- AWS. (s.f.). *¿Qué es el aprendizaje mediante refuerzo?* Obtenido de [https://aws.amazon.com/es/what-is/reinforcement-learning/#:~:text=Reinforcement%20learning%20\(RL\)%20is%20a,use%20to%20achieve%20their%20goals](https://aws.amazon.com/es/what-is/reinforcement-learning/#:~:text=Reinforcement%20learning%20(RL)%20is%20a,use%20to%20achieve%20their%20goals).
- Carrasco, A. J. (2020). *Estudio experimental de diversos algoritmos de aprendizaje por refuerzo*. MÁLAGA : UNIVERSIDAD DE MÁLAGA .
- CFBLOGS. (18 de Nov de 2023). *SDE Analysis in Inventory Management*. Obtenido de Cash Flow Inventory: <https://cashflowinventory.com/blog/sde-analysis-in-inventory-management/>
- Corporation, C. I. (17 de 08 de 2021). *Guía de CRISP-DM de IBM SPSS Modeler*. Obtenido de Conceptos básicos de ayuda de CRISP-DM: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>
- De Moor B.J., G. J. (2022). Reward shaping to improve the performance of deep reinforcement learning in perishable inventory management. *European Journal of Operational Research*, 535-545.
- Dehaybe, H., Catanzaro, D., & Chevalier, P. (2024). Deep Reinforcement Learning for inventory optimization with non-stationary uncertain demand. *European Journal of Operational Research*, 433-445.
- Fariborz Y, P., & Murugan, A. (2002). Classifying inventory using an artificial neural network approach. *Computers & Industrial Engineering*, 389-404.
- Guvenir, H., & Erel, E. (1998). Multicriteria inventory classification using a genetic algorithm. *European Journal of Operational Research*, 29-37.
- Kartal, H., Oztekin, A., Gunasekaran, A., & Cebi, F. (2016). An integrated decision analytic framework of machine learning with multi-criteria decision making for multi-attribute inventory classification. *Computers & Industrial Engineering*, 599-613.
- Kitchat, K., Lin, M.-H., Chen, H.-S., Sun, M.-T., K. S., Ku, W.-S., & Surasak, T. (2024). *A deep reinforcement learning system for the allocation of epidemic prevention materials based on DDPG*. Expert Systems with Applications.
- Matich, D. J. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones*. Rosario: Universidad Tecnológica Nacional – Facultad Regional Rosario .

- Navarro, M. P. (2022). *Aplicación de técnicas de aprendizaje por refuerzo a navegación visual*. Alicante: Universidad de Alicante.
- Rahman Mahin, M., Shahriar, M., Rani Das, R., Roy, A., & Wasif Reza, A. (2025). Enhancing Sustainable Supply Chain Forecasting Using Machine Learning for Sales Prediction. *Procedia Computer Science*, 470-479.
- Ramanathan, R. (2006). ABC inventory classification with multiple-criteria using weighted linear optimization. *Computers and Operations Research*, 695-700.
- Ravinder, H., & Misra, R. B. (2014). ABC Analysis For Inventory Management: Bridging The Gap Between Research. *American Journal Of Business Education* –, Volume 7, Number 3.
- Rolf, B., Jackson, I., Müller, M., Lang, S., Reggelin, T., & Ivanov, D. (2023). A review on reinforcement learning algorithms and applications in supply chain management. *International Journal of Production Research*, 7151-7179.
- Rolf, B., Jackson, I., Müller, M., Lang, S., Reggelin, T., & Ivanov, D. (2023). *A review on reinforcement learning algorithms and applications in supply chain management*. Berlín, Germany: INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH.
- Temizoz, T., Imdahl, C., Dijkman, R., Lamghari-Idrissi, D., & Van Jaarsveld, W. (2025). Deep Controlled Learning for Inventory Control. *European Journal of Operational Research*.
- Villegas-Ch, W., Maldonado Navarro, A., & Sanchez-Viteri, S. (2024). Optimization of inventory management through computer vision and machine learning technologies. *Intelligent Systems with Applications*, Volume 24.
- WMS, P. (2024). *FAQ: What is VED Analysis?* Obtenido de PULPO WMS: <https://blog.pulpowms.com/faq-what-is-ved-analysis-uncovering-the-importance-of-inventory-classification>
- Y. Chen, K. L. (2008). A case-based distance model for multiple criteria ABC analysis. *Computers & Operations Research*, 776-796.
- Zárate, L. G., & Lozada, M. Á. (2020). Beneficios de utilizar el Análisis ABC en la administración de inventarios en una Pequeña y Mediana Empresa (PyME) comercializadora en Tlaxcala, México. (*Benefits of use ABC Analysis in inventory management in a small business from Tlaxcala, México*).