

ANÁLISIS DE LOS ERRORES EN EL PICKING Y VERIFICACIÓN DE UNIDADES SUeltas EN EL CENTRO  
DE DISTRIBUCIÓN DE COMERCIAL NUTRESA BOGOTÁ

TRABAJO DE GRADO APLICADO

ENTREGA III

PRESENTADO POR:

DIEGO RAFAEL MILLÁN MARTÍNEZ

SANTIAGO ALEJANDRO TOVAR CAMPO

PONTIFICIA UNIVERSIDAD JAVERAINA

MAESTRÍA EN ANALÍTICA PARA LA INTELIGENCIA DE NEGOCIOS

FACTULDAD DE INGENIERÍA

BOGOTÁ D.C.

2019

## Contenido

Grupo Nutresa - Comercial Nutresa Bogotá .....	4
Entendimiento del Negocio.....	7
Antecedentes .....	7
Objetivos de negocio.....	10
Inventario de recursos .....	10
Requerimientos, supuestos y restricciones .....	11
Riesgos y contingencias.....	11
Terminología .....	12
Costos y beneficios.....	12
Objetivos de la minería de datos .....	13
Plan del proyecto .....	13
Evaluación inicial de herramientas y técnicas.....	14
Entendimiento de los datos .....	14
Reporte de la recolección inicial de los datos.....	14
Alistamientos consolidados.....	15
Novedades consolidadas.....	15
Usuarios.....	16
Reporte de descripción de los datos.....	16
Alistamientos consolidados.....	16
Novedades consolidadas.....	20
Usuarios.....	24
Reporte de calidad de los datos.....	24
Preparación de los datos.....	25
Identificación de variables preliminares .....	25
Construcción y exploración del dataset .....	26
Reporte de descripción del dataset .....	29
Modelos.....	33
Diseño del proceso .....	33
Descripción de modelos.....	33
Evaluación y métricas de modelos .....	33
Evaluación .....	35
Recomendaciones para el picking.....	35

Mejora de la verificación.....	35
Indicadores de evaluación.....	36
Modelo de Machine Learning .....	36
Máquina de control de peso .....	36
Análisis de propuestas.....	37
Anexos .....	39
Anexo 1.    Glosario de Terminología .....	39
Anexo 2.    Diagrama de procesos de picking.....	40
Anexo 3.    Sistema de vídeo para verificación de errores en picking.....	41
Bibliografía .....	42

## Grupo Nutresa - Comercial Nutresa Bogotá

“Grupo Nutresa S. A. es la empresa líder en alimentos procesados en Colombia y uno de los jugadores más relevantes del sector en América Latina. Fundada en 1920, cuenta en la actualidad con cerca de 45.600 colaboradores y opera a través de ocho unidades de negocio: Cárnicos, Galletas, Chocolates, Tresmontes Lucchetti -TMLUC-, Cafés, Alimentos al Consumidor, Helados y Pastas.” (Nutresa, Quiénes somos - Grupo Nutresa, 2019)

En Colombia, la venta y distribución de productos de las diferentes unidades de negocio son soportadas por tres empresas (Comercial Nutresa, Novaventa y La Recetta), donde cada una atiende un nicho de mercado diferente. (Cadenas de supermercado, Tradicional, Institucional, Industrial, Alimentos al consumidos, Alternativo). Comercial Nutresa es la empresa del grupo que se encarga de vender y distribuir los alimentos secos para el canal Tradicional, Autoservicios y las Grandes Superficies en Colombia.

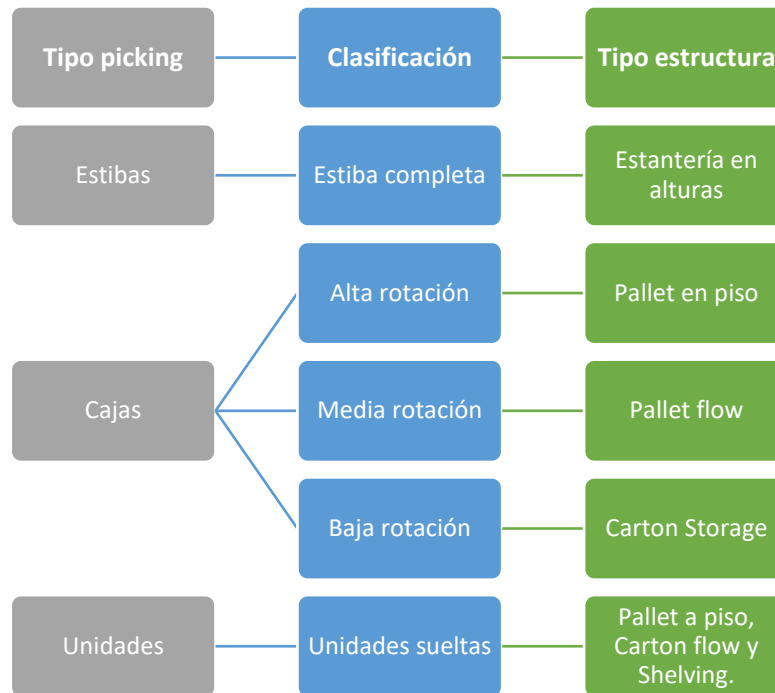


Ilustración 1 Estructura de Grupo Nutresa

De las ventas totales de Comercial Nutresa, la regional de Bogotá representa cerca del 33% de las ventas a nivel nacional, donde a su vez el 31% de las ventas de la regional se preparan desde el centro de distribución en Bogotá, ubicado en el Barrio Salazar Gómez.

En procura de mejorar el proceso de alistamiento y entrega de la mercancía a los clientes, el centro de distribución de la ciudad de Bogotá tiene un sistema mecanizado *pick to belt* que inició operación en enero del 2014, permitiendo realizar alistamientos de los pedidos individuales de cada cliente en el centro de distribución sin extensos recorridos, donde desde diferentes estaciones se alistan diferentes partes del pedido, siendo enviados a los muelles de carga con ayuda de bandas transportadoras para ser consolidados antes del cargue del vehículo. (Anaya Tejero, 2008; Anaya Tejero, 2008).

Cada uno de los pedidos se desagrega en sub almacenes de acuerdo a la cantidad de mercancía que se esté solicitando y el lugar donde se tenga parametrizado el SKU. La Ilustración 2 Tipos de sub-almacenes en el CEDI Bogotá – Mecanizado muestra los diferentes sub-almacenes del CEDI de Bogotá.



*Ilustración 2 Tipos de sub-almacenes en el CEDI Bogotá – Mecanizado*

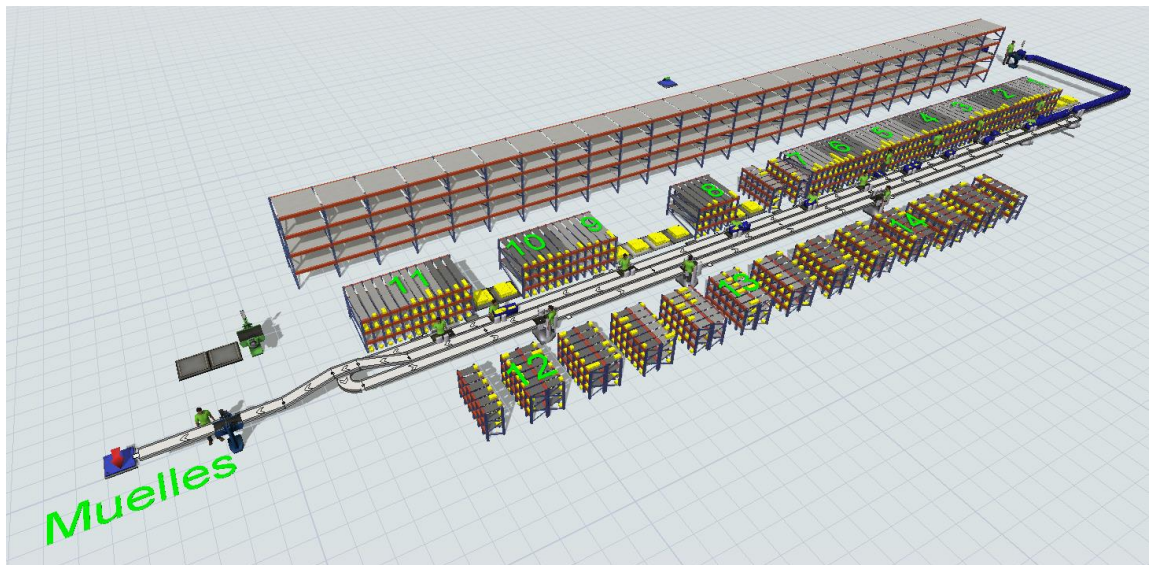
Es importante aclarar que cada material debe tener una ubicación de consumo de unidades sueltas en una única zona de alistamiento, debe tener una ubicación de cajas completas en un único subsistema (excluyentes), y puede tener varias ubicaciones en alturas de acuerdo al stock de seguridad.

A pesar de contar con diferentes sistemas, no todos los sistemas pasan por la banda transportadora y no todos los usuarios quedan registrados en el almacén. A continuación, algunas particularidades de cada tipo de picking:

- Estibas completas: Este almacén corresponde al almacenamiento general de mercancía, cuando la cantidad a almacenar corresponde a más de una estiba. Solo se alista de estas ubicaciones cuando un cliente solicita más de una estiba moviéndola con ayuda de un montacargas directamente a los muelles de cargue. Todos los alistamientos los confirma el montacarguista por medio de una radiofrecuencia.
- Cajas – Alta rotación: Cerca a los muelles hay unas ubicaciones en piso que corresponden a los materiales que se mueven con mucho volumen (sin necesidad de ser una estiba). Este alistamiento lo realiza un operario recorriendo las ubicaciones de la zona con un estibador, el cual registra sus alistamientos en el sistema con ayuda de una radio frecuencia. Luego de alistar la mercancía va al muelle de cargue correspondiente y deja la estiba con los productos alistados.
- Cajas – Media rotación: Ubicado en el segundo y tercer nivel del sistema mecanizado, este proceso es realizado por una persona que recibe todas las cajas a alistar de este subsistema para un transporte. Teniendo en cuenta que estas cajas se envían por el sistema mecanizado, el operario recibe una tira de etiquetas en las cuales el debe buscar la ubicación-posición donde está cada producto, colocar una etiqueta a cada una de las

cajas a alistar y subirlas a la banda transportadora para enviarla a los muelles. El operario no realiza ninguna confirmación en el sistema, ya que cuando pasa por el desviador de cajas sobre los muelles de cague, el *sorter* escanea la etiqueta para saber el muelle al cual debe enviarla y realiza la confirmación del alistamiento con un usuario genérico. Esto significa que no es posible saber qué usuario realizó el alistamiento si proviene de este tipo de *picking*.

- Cajas – Baja rotación: Los productos que son de baja rotación para el CEDI no son almacenados en estibas completas ya que su stock de seguridad es inferior a una estiba, siendo almacenados en *carton storage*. Al momento en que se solicita una caja de este producto, un montacargas especial para esta zona se encarga de realizar el alistamiento de acuerdo con la tira de etiquetas que le fue solicitado. Una vez termina de alistar todas las cajas solicitadas y coloca las etiquetas a cada una de ellas, las envía a los muelles por medio de la banda transportadora. Esto significa que no es posible saber qué usuario realizó el alistamiento si proviene de este tipo de *picking*.
- Unidades sueltas: El sistema de unidades sueltas (representado en la ilustración 3) es el corazón del sistema mecanizado, dado que es el que permite destellar desde el centro de distribución las unidades sueltas que pide cada uno de los clientes. Este sistema se compone actualmente de 14 zonas de picking con tres tipos de estructuras: Estiba completa, shelving y cartón Flow, como muestra la ilustración 4.



*Ilustración 3 Representación del sistema de picking de unidades sueltas*

El proceso realizado sobre este sistema inicia cuando al operario que realiza la inducción de contenedores se le da una tira de etiquetas para que le coloque cada una a un contenedor o caja. Cada etiqueta representa un serial el cual reúne algunos o todas las unidades sueltas de los pedidos del cliente, de forma que el contenedor va pasando de estación en estación hasta completar todos los materiales y unidades que debería llevar. Mientras el contenedor va viajando, en cada una de las estaciones el operario asignado va leyendo cada uno de los contenedores con una radio frecuencia, y esta muestra la orden de trabajo que se debe realizar. La pantalla puede mostrar de forma simultánea hasta dos

materiales solicitados, y en ese momento el operario debe desplazarse a la ubicación, leer el código de barras del producto para garantizar se está alistando lo que se debe, extraer la cantidad de unidades solicitadas, hacer la inducción de estas en el contenedor y confirmar la orden de trabajo en la radio frecuencia. Una vez termina, envía el contenedor a la siguiente estación y vuelve a leer otro contenedor.



Estiba completa



Shelving



Carton Flow

*Ilustración 4 Tipos de estructura en el sistema de picking de unidades sueltas.*

Independientemente del sub-almacén del que provenga la mercancía, una vez esta llega a los muelles de carga, se realiza una verificación de los productos alistados tanto en cajas como en unidades y se corrigen aquellos errores que fueron detectados. Esto se realiza con el fin de garantizar que la mercancía alistada corresponde a los pedidos solicitados y facturados para su entrega al cliente.

## Entendimiento del Negocio

Teniendo en cuenta la descripción del proceso general que se realiza en Comercial Nutresa Bogotá, se presenta a continuación el entendimiento del negocio y sus problemáticas actuales.

### Antecedentes

Para comercial Nutresa el proceso de *picking* dentro de su operación tiene dos conceptos fundamentales: Eficiencia y efectividad. La eficiencia dentro de la operación corresponde a la velocidad con la que realizan el alistamiento de los pedidos, mientras que la efectividad corresponde a la exactitud con que se alisto la mercancía que se había solicitado. Todo esto orientado a cumplir la promesa de servicio al cliente, la cual puede ser resumida en entregarle los productos solicitados, en el momento acordado y en las cantidades acordadas. A esto se le conoce como Nivel de Servicio.

*“La entrega certificada es una alianza en la que el proveedor y el cliente se comprometen a realizar todas las operaciones necesarias para garantizar al consumidor final total confiabilidad en términos de la calidad del producto, de su manejo físico, de la información que suministra y de la exactitud de los documentos que amparan cada transacción. No es más que entregar en las fechas y cantidades pactadas, materiales, productos y servicios que cumplan los requisitos de calidad exigidos por el cliente.” (Logística, 2018)*

Cuando la logística de la empresa no es capaz de cumplir esa promesa de valor, el inventario de los productos en el punto de venta corre el riesgo de agotarse y extender su periodo de reposición. Frente a esto, existen cuatro comportamientos predominantes del consumidor en la tienda que se resumen o conocen como venta perdida (Logyca, 2017):

- Compra otro producto de diferente marca.
- Compra otro producto de la misma marca, pero en presentaciones diferentes.
- Va a otro supermercado.
- No compra el producto.

Los causales más comunes son:

- Inexactitud en el inventario.
- Proveedor no entregó el producto.
- Producto no solicitado al proveedor.

*“Varias causales y paradigmas operativos caducos contribuyen a generar errores en la preparación de pedidos, lo que afecta el nivel de servicio a clientes, obliga a duplicar operaciones y gestiones administrativas y, en la búsqueda de responsabilidades, genera conflictos dentro de la compañía. Esta operación típica de los Centros de Distribución y bodegas, ya por el ambiente y los equipos de trabajo implicados, el diseño y método de trabajos, las tecnologías y calidades requeridas, puede en su deficiencia dañar seriamente los objetivos de una empresa en relación a la rentabilidad y en el vínculo de fidelidad con sus mercados. Las causas pueden ser internas (por la estructura, las características operativas o recursos aplicados) o externas... ambas deben ser detectadas y atacadas.”*  
(Demaría, 2003)

Cuando existen novedades en el proceso de alistamiento y no se corrigen de raíz o ajustan antes de enviar la mercancía al cliente, se corre el riesgo perder venta, aparte de dar la posibilidad de que el consumidor pruebe otro producto y genere algún tipo de afinidad sustituta por él. Adicionalmente, la promesa de valor hacia el cliente (tienda) se ha roto, sabiendo que en algunas ocasiones sus clientes se dirigirán a otra tienda ya que no tiene el producto que estaba buscando. Esto empieza a generar inconformidad sobre el cliente, a tal punto de afectar la imagen de marca y la reputación de su proveedor, de rechazar pedidos o hasta dejar de comprarle.

En Comercial Nutresa las operaciones logísticas se evalúan con base tres pilares estratégicos con indicadores individuales. En cuanto a la logística de almacenamiento, los indicadores que impacta de forma directa son:

- Eje - Crecimiento rentable:
  - Cumplimiento del presupuesto de gastos: Cumplimiento del gasto presupuestado por el área para el año actual. Este indicador impacta el costo por servir y afecta la rentabilidad de la compañía.
- Eje - Organización centrada en el cliente:



- Facturado No Entregado: Mercancía alistada, facturada y enviada al cliente, pero que este no recibe por diferentes motivos. Este evento impacta el cumplimiento del presupuesto de ventas y el nivel de servicio.
- Eje - Capacidades Organizacionales:
  - Exactitud del inventario: Evitar la pérdida/sobrante de mercancía en el centro de distribución.

Actualmente, Comercial Nutresa presenta constantemente novedades en el producto que le envía a sus clientes, a pesar de realizar un esfuerzo enorme por tratar de identificar todas las novedades antes de enviar la mercancía. Algunos de los síntomas más importantes son los siguientes:

- Entre los meses de noviembre 2018, diciembre 2018 y enero de 2019, el 12.6% de los pedidos que fueron facturados y no entregados, estuvieron bajo el causal de *error en despachos*.
- Para el 2019 se proyecta gastar cerca de 480 millones en el personal requerido y exclusivo para hacer la verificación. Al mes de marzo, este concepto del presupuesto presenta una sobre ejecución cercana al 80% con tendencia a empeorar por aumento del personal.
- En los últimos tres meses el nivel de confiabilidad de los inventarios de las ubicaciones de consumo estuvo alrededor de un 30%.
- De todos los errores identificados, el 60% fue verificado sin novedad cuando si existía. Esto cuestiona la efectividad del proceso de verificación actual.
- Al colocar este proceso en los muelles, el proceso de cargue del vehículo se demora cerca de 30 minutos adicionales mientras se termina de verificar la mercancía, aunque ya este “completa” para el cargue.

El Número de errores con relación a la cantidad de alistamientos esta por alrededor de un 1%. Esto no representa mayor cantidad de novedades, pero si representa un malestar para el cliente, haciendo que lo devuelva en algunas ocasiones, en otras lo recibe y se le hace una nota crédito que repercute en las ventas, y se están destinando cerca de 500 millones de pesos al año para controlar este proceso.

Otro de los retos de mejorar la calidad del alistamiento y de la verificación radica en dos aspectos principales:

- Personal rotativo:
 

Aproximadamente el 80% del personal de *picking* es personal contratado por medio de empresas temporales. Este porcentaje se mantiene con el tiempo debido a las políticas de contratación actuales para este tipo de personal, haciendo que la empresa tenga curvas de aprendizaje de forma constante.
- Depuración e introducción de nuevos materiales:
 

A lo largo del año se presentan diferentes temporadas y lanzamientos de nuevos productos que pueden estar por tiempo limitado o que pueden agregarse al portafolio de ventas. Esto implica que a los operarios se les están cambiando con regularidad algunos productos, y el desconocimiento de estos pueden generar novedades en el *picking*.

Teniendo en cuenta la complejidad del proceso, se desea aplicar herramientas analíticas que permitan ayudar a mejorar el error generado en el alistamiento de unidades sueltas, ya que cerca

del 95% de las novedades ocurren por diferencias de conteo al hacer el alistamiento. El 5% restante corresponde a novedades que se presentan en cajas completas, caracterizadas por cruzar materiales (Se solicitaba pasta y se alisto un producto de chocolates), pero no presenta problemas de cantidad dado que se alista con base en las etiquetas impresas y entregadas al operario.

### Objetivos de negocio

Los objetivos están son disminuir la cantidad de novedades que se presentan en el alistamiento y requerir la menor cantidad de mano de obra para verificarlo.

- Identificar al menos 2 recomendaciones clave para disminuir el número de novedades por error en el picking, señalando patrones y características comunes en los alistamientos problemáticos que pueden ser resueltas antes de la verificación.
- Reducir la verificación manual del 100% de los alistamientos utilizando un modelo predictivo para identificar los contenedores y materiales que tienen un perfil de novedad. La verificación inteligente a través de la clasificación de los alistamientos tendrá impacto en la operación de Comercial Nutresa, haciéndola más rentable al disminuir los costos asociados al proceso.

### Inventario de recursos

Los recursos disponibles para el proyecto son los siguientes:

- Recursos de Hardware:
  - Propios: El equipo de trabajo cuenta con dos computadores personales con características de procesador, memoria RAM, tarjeta gráfica avanzadas. Estos dispositivos tienen el software necesario para el desarrollo del proyecto.
  - Comercial Nutresa: Todos los empleados administrativos de la empresa cuentan con computador Dell o Lenovo de buenas características (Procesador i5 o i7 de 7ma generación y 8 Gb de memoria RAM).
- Recursos de datos:
  - Del sistema SAP:
    - Bases de datos de todos los alistamientos realizados en los últimos tres meses en la operación de Bogotá OP.
    - Bases de datos de devoluciones de mercancía al final de la ruta de entrega por errores de alistamiento.
  - Datos de registro manual:
    - Consolidado de novedades pertenecientes al almacén.
      - Seguimiento de novedades identificadas en los muelles de cargue.
      - Seguimiento de novedades reportadas como faltantes por empresas de transporte.
    - Creación de nuevos usuarios en SAP.
    - Capacitación de personal en laboratorio simulado XLab.
  - Documentación escrita:
    - Documentos de consultoría y estudios pasados sobre el sistema de alistamiento de Comercial Nutresa Bogotá.

- Recursos de persona:
  - Sponsor del proyecto: jefe de almacén y transporte regional Bogotá.
  - Administradores de los sistemas de información y expertos del proceso en Comercial Nutresa Bogotá:
    - Analista de Slotting de Comercial Nutresa Bogotá.
    - Analista de WMS de Comercial Nutresa Bogotá.
    - Auxiliar para seguimiento de mercancía faltante en ruta.
    - Coordinadores de operación bajo techo.
  - Científicos de datos:
    - Estudiantes de la javeriana que presentan el proyecto/consultoría.
    - No hay expertos en estadística y sistemas en el equipo de logística de Comercial Nutresa.

### Requerimientos, supuestos y restricciones

A continuación, se enlistan los requerimientos, supuestos y restricciones del proyecto:

- Requerimientos:
  - El proyecto va a estar acotado a la operación del Centro de Distribución de la Ciudad de Bogotá.
  - La información suministrada por Comercial Nutresa será compartida a partir de su correo corporativo google drive. Se tiene permiso para hacer uso de su información para el desarrollo del proyecto.
  - Tener en cuenta que el almacén está constantemente actualizando sus SKUs, dado que hay productos que se venden por tiempo limitado.
  - Tener en cuenta que una parte del personal rota de forma constante.
- Supuestos:
  - Se asume que los datos suministrados por Comercial Nutresa son reales y no tienen problemas desde su captura.
  - La verificación actual de los alistamientos se está realizando al 100%.
  - Se asume que se tienen identificadas todas las novedades del almacén desde el reporte que resume estas eventualidades.
  - Se asume que los datos están actualizados antes de ser entregados.
- Restricciones:
  - El desarrollo del proyecto tiene que llevarse a cabo en el primer semestre académico del 2019.
  - No se permite la copia, divulgación y/o publicación de los datos suministrados por Comercial Nutresa.
  - Las reuniones de entendimiento y avance con Comercial Nutresa están sujetas a la disponibilidad de tiempo del equipo.

### Riesgos y contingencias

Algunos de los riesgos identificados son:

- Cambio de jefe: En caso de que la persona encargada del proyecto no pueda seguir haciéndose cargo del proyecto, la Directora de Logística de Bogotá será la nueva responsable.

- Demora en el suministro de información: Para facilitar el acceso a la información, se estableció que hay dos personas que nos pueden dar respuesta a las solicitudes: Analista de gestión logística y Analista de WMS.
- Posible pérdida de información: Se establece como medio de transferencia de información entre los estudiantes y la empresa a Google Drive, ya que las cuentas de correos corporativos están soportadas en estos sistemas, lo que permite tener un respaldo de la información compartida.
- Problemas con la calidad de los datos: En caso de encontrar algunos datos incongruentes desde la información suministrada, se realiza revisión inmediata con las personas asignadas para el suministro de información.

## Terminología

La terminología básica se encuentra en el Glosario de Terminología.

## Costos y beneficios

A continuación, se muestran algunos análisis costos y beneficios que pueden presentarse al final del proyecto:

- Posibles costos para incurrir en la aplicación del proyecto:
  - No se incurre en costos por la recolección de datos, dado que Comercial Nutresa ya cuenta con sistemas y personas responsables para este proceso como parte de su rutina diaria.
  - Teniendo en cuenta que es posible que del proyecto se identifiquen algunas herramientas útiles para diferentes procesos del almacén (Ej: Análisis de materiales para *slotting*), sería necesario pagar por una adaptación para desarrollarlo como una herramienta replicable para alguien que no tiene los conocimientos estadísticos y de programación que podrían ser requeridos.
  - Si el desarrollo de una solución se realiza sobre uno de los computadores, no es necesario incurrir en ningún gasto adicional, pero si se considera desarrollar un software local o web, sería necesario realizar un desarrollo adicional y cotizarlo.
  - Si se recomienda el uso de otras herramientas que contribuyan a disminuir las novedades (Ej: control de peso), sería un costo adicional en el cual se podría incurrir.
- Posibles beneficios:
  - Actualmente se proyectan pagar cerca de 480 millones de pesos por el personal que está exclusivamente dedicado al proceso de verificación (Mojica, 2019). Si se logra realizar una verificación predictiva se generarían ahorros que contribuyan al costo por servir y a la rentabilidad de la compañía. En este caso el ROI dependerá exclusivamente de la cantidad de recursos que se puedan prescindir y de la inversión para desarrollar la herramienta.
  - Aumentar la confiabilidad de los inventarios de consumo por errores en los alistamientos.
  - Disminuir el valor de pedidos no recibidos por errores en despacho.
  - Generar ahorros por no realizar impresión de las listas de chequeo por verificación.

- Disminuir las notas crédito aplicadas al cliente por entregar mercancía incompleta. Este representa ventas perdidas y no contamos con información por parte de la empresa.

## Objetivos de la minería de datos

Para dar cumplimiento a los objetivos planteados con una perspectiva de negocio,

- Identificar al menos 2 recomendaciones clave para disminuir el número de novedades por error en el picking:
  - Este objetivo puede ser abordado realizando un análisis detallado de la información suministrada, soportado en modelos no supervisados, con el fin de identificar las diferencias significativas de las variables a evaluar sobre el error en el alistamiento.
- Reducir la verificación manual del 100% de los alistamientos utilizando un modelo predictivo para identificar los contenedores que tienen un perfil de novedad:
  - Este objetivo corresponde a un problema de clasificación, en el cual se desea identificar los alistamientos que son candidatos para tener novedad. El modelo debería identificar más del 40% del alistamiento con novedades sin la necesidad de revisar el 100% de los contenedores.

## Plan del proyecto

El plan de trabajo del proyecto se construyó incluyendo una etapa llamada inicio del proyecto, el cual consistía crear compromiso con el jefe del almacén en el proyecto, realizar una visita a las instalaciones del centro de distribución para comprender el proceso, tener una reunión con los coordinadores de operación para que manifestaran sus inconformidades y oportunidades con la calidad del alistamiento del sistema de picking e identificar la información disponible para el desarrollo del modelo. Las fases posteriores corresponden a las fases sugeridas por la metodología CRISP-DM con relación a las fechas de entrega acordadas con la universidad.

Etapas	Actividades	Entrega I				Entrega II				Entrega final								
		4-feb	11-feb	18-feb	25-feb	4-mar	11-mar	18-mar	25-mar	1-abr	8-abr	15-abr	22-abr	29-abr	6-may	13-may	20-may	27-may
Inicio del proyecto	Reunión de iniciación con jefe de almacén	■																
	Acompañamiento a la operación del CDH		■															
	Reunión con coordinadores de operación			■														
Business understanding	Conocer data disponible y solicitarla				■													
	Determine business objectives					■												
	Asses situation						■											
Data understanding	Determine data mining goals						■											
	Produce project plan							■										
	Collect initial data								■									
Preparar entrega I	Describe data									■								
	Explore data										■							
	Verify data quality											■						
Data preparation	Documento																	
	Presentación																	
	Select data																	
Modeling	Clean data																	
	Construct data																	
	Integrate data																	
Preparar entrega II	Format data																	
	Dataset																	
	Select modeling techniques																	
Evaluation	Generate test design																	
	Build model																	
	Asses model																	
Deployment	Documento																	
	Presentación																	
	Evaluate results																	
Preparar entrega final	Review process																	
	Determine next steps																	
	Plan deployment																	
Preparar entrega final	Plan monitoring and maintenance																	
	Produce final report																	
	Review project																	
Preparar entrega final	Documento																	
	Presentación																	

Ilustración 5 Plan de trabajo del proyecto

A medida que se avanza en el proyecto, es probable que se vuelva a alguna etapa preliminar con el fin de solicitar nueva información o validar algo de la que la había sido suministrada. Esto aplica para el entendimiento del negocio hasta de las bases de datos.

### Evaluación inicial de herramientas y técnicas

Se utilizará Jupyter-Notebook, un ambiente especial con herramientas optimizadas para aplicar analítica, construidas usando el lenguaje de programación Python. Jupyter incluye los siguientes paquetes código abierto:

- Pandas: librería de software que provee estructuras de datos de alto rendimientos, además de herramientas para análisis de datos.
- Tensorflow: librería de software para desarrollar y entrenar modelos de machine learning.
- Keras: librería de software para aplicar deep learning.
- Seaborn: librería de software para visualización de datos.
- Matplotlib: librería de software para visualización de datos.
- Numpy: librería de software para realizar cálculos científicos.

En cuanto a las técnicas a utilizar en el proyecto, se pueden clasificar en dos tipos de acuerdo con los objetivos planteados:

- Predictivas: Clasificación
- Descriptivas: Agrupamiento, exploración y visualización de datos

### Entendimiento de los datos

El entendimiento de los datos se realizó para cada una de las bases de datos que fueron suministradas para el desarrollo del proyecto:

- Alistamientos consolidados
- Novedades consolidadas
- Usuarios

Todos los archivos fueron suministrados en Excel con algunos campos nombrados de forma diferente, por lo que antes de realizar todo el proceso de entendimiento de datos se consolidaron en Python, de forma que pudiéramos tener una sola base de datos manipulable con relación a los requerimientos y objetivos del proyecto.

La información suministrada fue de los meses de noviembre de 2018, diciembre de 2018 y enero de 2019.

### Reporte de la recolección inicial de los datos

Con base en el listado de bases de datos suministradas por Comercial Nutresa, se realiza la descripción de cada una de ellas aclarando el método de extracción y/o construcción de los datos.

### Alistamientos consolidados

La información de los alistamientos consolidados es consultada y descargada de SAP. Durante el proceso de picking, cada vez que una orden de trabajo (OT) es confirmada, el sistema crea un registro con la información de la cantidad teórica, la cantidad alistada, el usuario que la alistó, la hora en que fue alistada, la ubicación del que hizo la extracción de la mercancía, entre otros.

### Novedades consolidadas

La base de datos suministrada resume las novedades identificadas en el proceso de picking del almacén. Esta se construye de forma manual y tiene diferentes fuentes de información y procesos para su consolidación. Está compuesta de los siguientes elementos:

#### *Novedades identificadas y reportadas en la verificación antes de cargar los vehículos:*

El proceso de verificación en los muelles se realiza con hojas impresas con todo lo que está esperando por recibir en los muelles.

#### *Novedades reportadas en la mañana por ayudante de entrega de TAT:*

Los ayudantes de venta en la mañana revisan sus vehículos y en caso de encontrar una diferencia, diligencian un formato con las novedades y van al equipo de inventarios para que revisen y complementen el faltante.

#### *Novedades luego de la ruta de entrega:*

Al regresar los vehículos al CEDI, estos realizan una legalización de todos los viajes completados y de las novedades de los pedidos que tenían que entregar. Entre esas novedades, devuelven la mercancía que estaba sobrando y legalizan las notas crédito que se deben procesar porque la mercancía no llegó completa (en caso de que el cliente reciba el pedido).

Teniendo en cuenta que no se sabe si el error de la mercancía se generó por el picking o por algún evento en el transporte, existe una persona bajo la figura de auxiliar de operaciones que se encarga de hacerle seguimiento a las pérdidas de mercancía siguiente el siguiente protocolo:

1. Organizar las novedades de faltantes en ruta de mayor a menor impacto.
2. Revisar los alistamientos para identificar la hora del alistamiento y el lugar.
3. Revisa videos del proceso del alistamiento para validar si existió alguna novedad.  
Anexo 3
4. Revisa los listados de verificación diligenciados por el personal encargado, con el fin de evidenciar si se realizó una corrección del alistamiento o si se había validado como correcto.
5. Revisa en cámaras el proceso de verificación.

Luego de hacer todo el seguimiento, hay tres posibilidades:

1. Existió un error en el alistamiento y por la velocidad requerida de la operación no se validó: Error en el alistamiento.
2. Existió un error en el alistamiento, se validó, pero el verificador no cayó en cuenta de la novedad: Error en verificación.

3. Existe evidencia suficiente de que el producto se alisto bien, se verifico y cargo de forma adecuada, por lo que se cobra la perdida de mercancía al transportista.

Teniendo en cuenta esté procedimiento, la base de datos de las novedades consolidadas corresponde al documento más confiable que tiene Comercial Nutresa para identificar los errores de alistamiento existentes en el proceso de *picking*.

## Usuarios

La información del listado de los usuarios es construida de forma manual por la persona responsable de solicitar todos los usuarios del personal nuevo. Está persona registra en un campo los nombres completos de cada uno de los operarios y en otro la fecha de activación del usuario. Este dato puede ser usado como la fecha de ingreso de la persona.

La misma persona está encargada de hacer seguimiento del proceso de capacitación que se realiza al personal nuevo y antiguo, guardando los resultados de la prueba en formato de imagen, pero realizando un registro de que esa persona participo en el proceso de formación.

El proceso de formación se realiza en un simulador físico con características similares a las que se encuentra en las zonas de alistamiento de unidades sueltas, utilizando productos reales y ordenes de pedido ficticias.

## Reporte de descripción de los datos

A continuación, se presenta la descripción de cada una de las bases de datos suministradas por Comercial Nutresa:

### Alistamientos consolidados

La base de datos de los alistamientos consolidados pesa 466 Mb y tiene , las cuales se describen a continuación.

Variable	Tipo	Descripción
OT	Id	Numeración de las ordenes de trabajo generadas por el sistema. Esta solicita uno o más alistamientos y debe ser ejecutada por una única persona en una zona de trabajo.
Pos.	Numérico	Identificación del número del alistamiento a realizar dentro de la OT.
Conf.	Bin	Variable Binarias {X,""} para identificar qué registro fue confirmado y cuál no.
Transporte	id	Numeración del transporte donde se agrupan los pedidos de los clientes.
Entrega	id	Homólogo del número de pedido, pero para el almacén.
SEGMENTO	Texto	Canal al cual corresponde al pedido.
Muelle	Texto	Muelle de cargue en el que se encuentra el vehículo que va a realizar la entrega.
Serial	Texto	Identificador del contenedor o caja que se alista y pasa por el sistema mecanizado. Es requerido para que sea leído por un sorter y sea desviado al muelle de cargue que corresponde e manera automática.
Caja	Texto	Numeración de los corrugados utilizados.



Variable	Tipo	Descripción
Secuenciadefactura	id	Orden para la secuencia de facturación a generar
Mat.Embalaje	Texto	Tamaño de la caja si el alistamiento se realiza del almacén de unidades sueltas (401).
TAP	id	Código del almacén del cual se realiza el alistamiento (401: Unidades Seltas, 402: Caja completa segundo/tercer nivel mecanizado, 100: Caja completa carton storage, 405: Caja completa ubicaciones de piso - masivos, 121: Estiba completa de alturas, Otros)
ÁreaPicking	Texto	En el caso del almacén de unidades sueltas (401), aclara el área de picking del cuál se realizó el alistamiento.
Ubicación	Texto	Identificador de la ubicación de la cual se realizó el alistamiento. Puede corresponder a pasillo, columna, nivel.
Material	id	SKU o código que identifica al producto dentro del almacén.
Descripción	Texto	Descripción del producto solicitado.
Lote	Texto	Número identificador del lote de producción del material
FechaCaducidad	Fecha	Fecha de vencimiento del producto.
Centro	Id	Código identificador del negocio dentro del almacén. (Chocolates, cárnicos, café, pasta, galletas, otros)
Almacén	Texto	Codificación del tipo de negocio con relación al almacén
Cant.Teórica	Numérico	Cantidad solicitada de acuerdo con el pedido realizado por el ejecutivo comercial.
Cant.Separada	Numérico	Cantidad alistada y confirmada por el operario. En algunos casos puede sr inferior si no existe suficiente mercancía.
Unid.	Cadena	Unidad de medida que se está solicitando alistar. Unidades es la más común para todos los alistamientos.
Ind.Diferencia	Bin	Variable Binarias {X,""} para identificar qué no se alisto la cantidad solicitada.
PesoAlistado	Numérico	Valor del peso alistado y confirmado por el operario.
UnidaddePeso	Texto	Unidad de medida del peso alistado.
VolumenAlistado	Numérico	Valor del peso alistado y confirmado por el operario.
UnidaddeVolumen	Texto	Unidad de medida del peso alistado.
Usuario	Texto	Código del usuario que realizo el alistamiento.
Fecha	Fecha	Fecha en la cual se realizó el alistamiento
Hora	Hora	Hora en la cual se realizó el alistamiento.

*Tabla 1 Descripción de datos Alistamientos consolidados*

A continuación, se presenta la tabla resumen de las variables numéricas de la base de datos:

	Cant.Teórica	Cant.Separada	PesoAlistado	VolumenAlistado
<b>count</b>	1,797,072.00	1,797,072.00	1,797,072.00	1,797,072.00
<b>mean</b>	14.04	13.73	4.98	0.0140
<b>std</b>	84.54	83.60	29.33	0.0873
<b>min</b>	-	-	-	-
<b>25%</b>	1.00	1.00	0.61	0.0020
<b>50%</b>	4.00	4.00	1.26	0.0040
<b>75%</b>	12.00	12.00	3.64	0.0090
<b>max</b>	21,120.00	21,120.00	6,096.38	41.06

*Tabla 2 Resumen alistamientos consolidados*

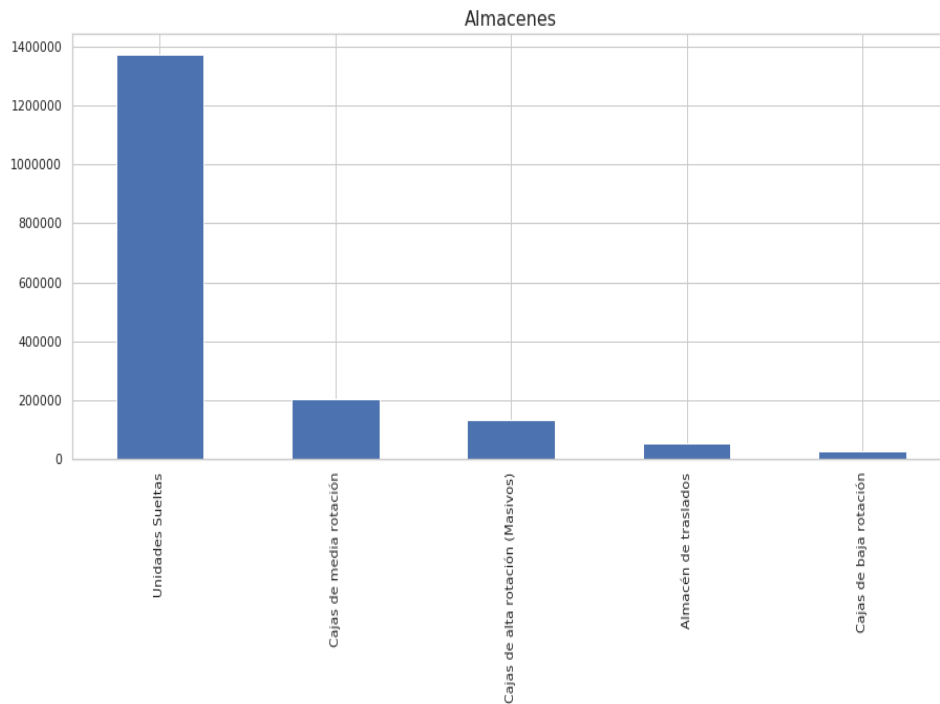
En el caso de los valores con NAs, se evidencia que existen 263 registros totalmente en blanco. Los demás vacíos son naturales teniendo en cuenta las características de cada sub-almacén, ya que no todos utilizan códigos de seriales para ser desviados por el *sorter*, ni requieren de un corrugado como medio de transporte.

<b>Variable</b>	<b># Nas</b>
OT	263
Pos.	263
Conf.	29079
Transporte	2166
Entrega	263
SEGMENTO	20500
Muelle	72910
Serial	191301
Caja	191632
Secuenciadefactura	191794
Mat.Embalaje	425176
TAP	263
ÁreaPicking	188668
Ubicación	263
Material	263
Descripción	263
Lote	263
FechaCaducidad	1271
Centro	263
Almacén	263
Cant.Teórica	263
Cant.Separada	263
Unid.	263
Ind.Diferencia	1794722
PesoAlistado	263
UnidaddePeso	263
VolumenAlistado	263
UnidaddeVolumen	263
Usuario	29079
Fecha	29079
Hora	263

*Tabla 3 Resumen de NAs Alistamientos consolidados*

En general, la gran parte de los alistamientos están relacionados con el código de almacén 401, que corresponde al sitio donde se alistan unidades sueltas (Tal como fue aclarado en la sección de descripción de datos). Esto se puede evidenciar en la ilustración 6, de donde se deduce que alrededor del 75% caen en la categoría mencionada anteriormente.

- 401: *Picking* de unidades sueltas.
- 402: *Picking* de cajas de media rotación.
- 405: *Picking* de cajas de alta rotación (Masivos).
- 200: Almacén de traslados.
- 100: *Picking* de cajas de baja rotación.
- 121: Almacenamiento y *picking* de estibas completas.



*Ilustración 6 Alistamientos por almacén.*

De forma similar, los alistamientos ocurren en diferentes áreas de picking. En la Ilustración 7 Alistamientos por área de picking, se pueden apreciar las áreas utilizadas, junto con la cantidad de alistamientos que ocurren en ellas. Dichas zonas de picking son usadas en proporciones similares, a excepción de las áreas 1 y 5, en las cuales ocurren alrededor del 21.6 % de los alistamientos. Lo anterior sugiere que hay una carga excesiva en ambas zonas.

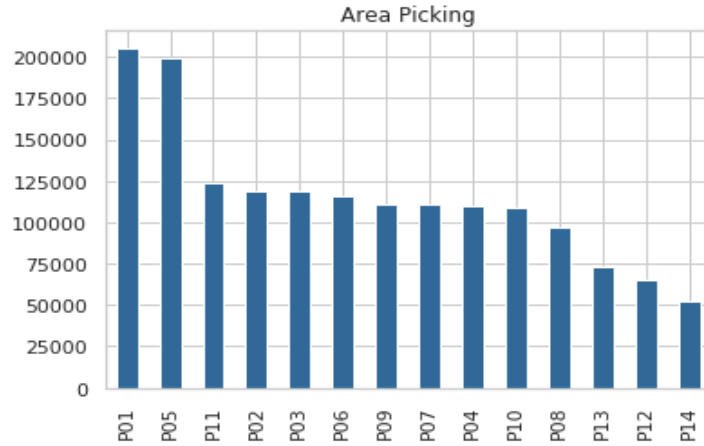


Ilustración 7 Alistamientos por área de picking

En la ilustración 7, se observan los canales para los cuales se alistan las órdenes, en conjunto con su respectiva proporción. De esos, los segmentos “AUTOBOGOTA”, “TAT” y “ZIPAQUIRÁ”, representan alrededor del 65% de los alistamientos realizados.

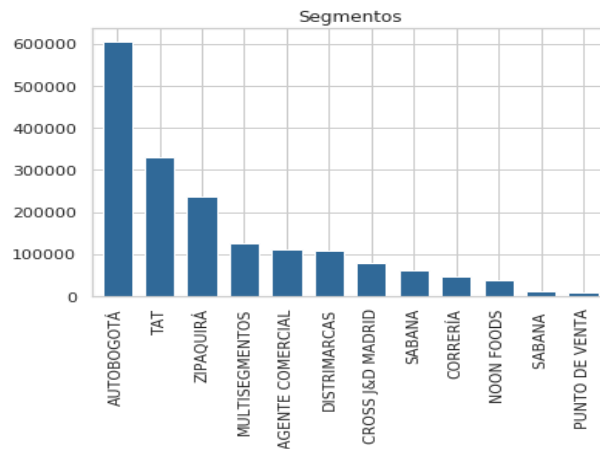


Ilustración 8 Cantidad de alistamientos por segmento

### Novedades consolidadas

La base de datos de novedades consolidadas pesa 1.2 Mb y tiene 11.411 registros con 10 variables, las cuales se describen a continuación.

Variable	Tipo	Descripción
FECHANOVEDAD	Fecha	Fecha en la que fue reportada la novedad. En algunos casos el ingreso de la novedad al sistema puede corresponder al día siguiente al alistamiento.
TRANSPORTE	id	Numeración del transporte donde se agrupan los pedidos de los clientes.
ENTREGA	id	Homólogo del número de pedido, pero para el almacén.
SERIAL	id	Identificador del contenedor o caja que se alista y pasa por el sistema mecanizado. Es requerido para que sea leído por un <i>sorter</i> y sea desviado al muelle de carga que corresponde y manera automática.
NOVEDAD	Texto	Tipo de novedad. Faltante, sobrante o avería por la manipulación.

MATERIAL	id	SKU o código que identifica al producto dentro del almacén.
CANTIDAD	Texto	Muelle de cargue en el que se encuentra el vehículo que va a realizar la entrega.
UN	Numérico	Cantidad de unidades que presentan novedad en el alistamiento.
CAUSALNOVEDAD	Texto	Descripción del motivo por el cual se presenta la novedad. Error en alistamiento o error en verificación.
REPORTEDE	id	Fuente de datos desde la que se identificó el reporte.

*Tabla 4 Descripción de datos Novedades consolidadas*

A continuación, se presenta la tabla resumen de las variables numéricas de la base de datos:

CANTIDAD	
<b>count</b>	11403
<b>mean</b>	4.238139
<b>std</b>	15.909002
<b>min</b>	0.5
<b>25%</b>	1
<b>50%</b>	1
<b>75%</b>	3
<b>max</b>	960

*Tabla 5 Resumen de novedades consolidadas*

La base de datos de novedades presentar varios NAs, principalmente en el número de la entrega y el causal de novedad. Sin embargo, para efectos del proyecto, los campos clave que permiten realizar el cruce de la información con la base de datos de alistamiento son el serial y el material. En este caso se asume que los que no tienen serial, no corresponden a alistamientos provenientes del 401.

Variable	# Nas
FECHANOVEDAD	5
TRANSPORTE	99
ENTREGA	5973
SERIAL	2035
NOVEDAD	23
MATERIAL	7
CANTIDAD	8
UN	12
CAUSALNOVEDAD	4211
REPORTEDE	1212

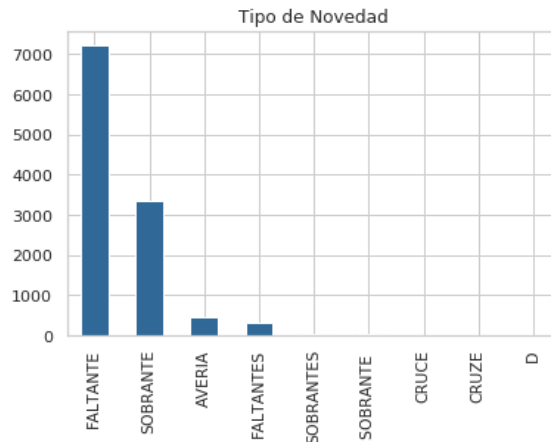
*Tabla 6 # Nas Novedades consolidadas*

Las novedades registradas en la base de datos están concentradas en cuatro tipos principales:

- Faltante: Se envió una cantidad inferior a la solicitada.

- Sobrante: Se envió una cantidad superior a la solicitada.
- Avería: La mercancía se alisto, pero presento daño en el producto haciendo necesario hacer un cambio.
- Cruce: Corresponde a hacer el alistamiento en las cantidades correctas, pero del material equivocado. En muy pocas ocasiones se logra identificar, ya que puede ser visto como el faltante de un material y el sobrante de otro.

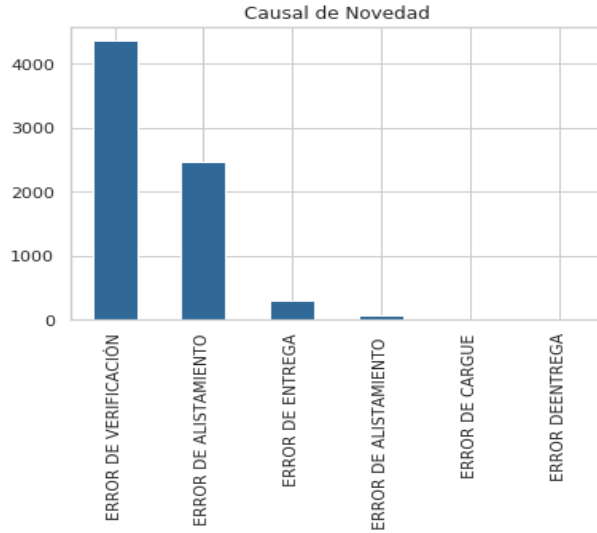
A pesar de que la base de datos presenta diferencias en las categorías asignadas, no es necesario hacer un ajuste de estas dado que todas representan una novedad en el alistamiento.



*Ilustración 9 Partición por tipo de novedad.*

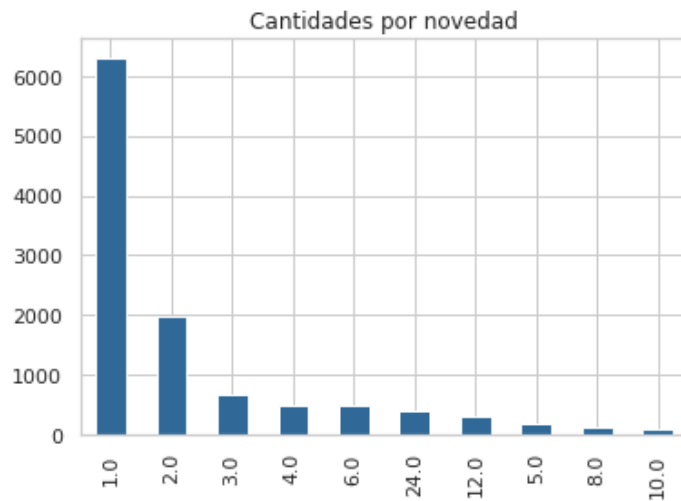
La base de datos de novedades identifica las causales de novedad:

- Error de alistamiento: Se identifico un error en el alistamiento que fue reportado por el verificador o reportado después del proceso de entrega sobre la mercancía que no se verificó (validado por la persona responsable del seguimiento de las novedades después de ruta).
- Error de verificación: Sucede cuando la mercancía alistada tenía alguna novedad, pero el verificador realizo una confirmación como si el alistamiento hubiera sido conforme.
- Error de entrega: Confusión sobre el vehículo al que se cargó la mercancía.



*Ilustración 10 Causal de novedad*

Revisando la cantidad de unidades que son reportadas con novedad, el 73% de las novedades corresponden a 1 o 2 unidades, lo que representa que las desviaciones pequeñas con relación a la cantidad solicitada. Esto se puede deber a un error en el proceso de conteo al momento de extraer el producto de la estantería.



*Ilustración 11 Cantidades reportadas con novedad.*

El 80% de las novedades están siendo generadas por alrededor de 200 materiales, que corresponden al 26% de los SKUs existentes en el almacén. A pesar de que existe una proporción pareto de 80/26, operativamente es difícil ejercer control sobre 200 materiales, teniendo en cuenta las diferencias físicas de cada uno de ellos.

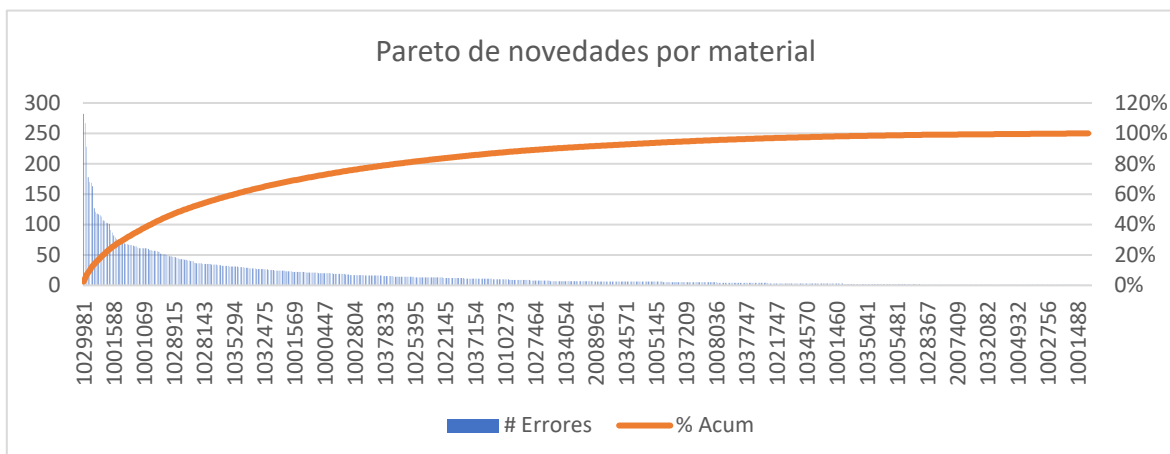


Ilustración 12 Pareto de novedades por material

## Usuarios

La base de datos de usuarios pesa 8 Kb tiene 136 registros con 5 variables:

Variable	Tipo	Descripción
Usuario	Texto	Código del usuario que realizó el alistamiento. Cada operario de picking tiene su usuario en el sistema.
FECHA INGRESO	Fecha	Fecha en que se creó el usuario en el sistema.
NOMBRES	Texto	Nombres del operario
APELLIDOS	Texto	Apellidos del operario
XLAB	Bin	Variable binaria que determina si la persona paso por un entrenamiento en un sistema simulado antes de iniciar sus labores en el proceso de picking.

Tabla 7 Descripción de datos Usuarios

Por el momento, las bases de datos de forma independiente están caracterizadas por tener varios identificadores (lo cual tiene sentido bajo la orientación de control del WMS), por lo que la exploración de los datos está orientada en las variables numéricas y categóricas principalmente.

## Reporte de calidad de los datos

El procedimiento para garantizar la calidad de los datos se aplicó de la siguiente forma:

- Tratamiento de valores nulos: fueron reemplazados por valores coherentes en la medida que el set de datos lo permitía; en caso contrario se elimina el registro de la base de datos pues podría afectar el resultado de los estudios posteriores.
- Modificación de registros con errores: algunos registros afectan la integridad de los datos pues carecen de confiabilidad. En aquellos casos donde se pudo detectar problemas relacionados a errores humanos al cargar datos, se modificaron los registros. Por ejemplo, entre los tipos de novedades se encontraron dos valores que representan lo mismo



("sobrante" y "sobrantes"); se asume que hubo un error y por tanto se colocó el mismo valor para cada registro ("sobrante", en este caso).

- Filtro de registros: Los registros fueron filtrados de tal forma que, solo aquellos alistamientos del tipo unidades sueltas, y cuyos canales de venta son principales, fueron seleccionados para el estudio en cuestión. Dichos registros son los más importantes pues representan casi la totalidad de las incidencias o novedades reportadas.

## Preparación de los datos

### Identificación de variables preliminares

Para el desarrollo de este proyecto, se han definido tres categorías de causas influyentes en el error del proceso de alistamiento (Demaría, 2003):

#### *Características de la persona*

Una de las variables que pueden influir en la generación de errores en el proceso de picking, corresponde al perfil, capacidad y experiencia de cada una de las personas.

Estudios previos han demostrado que el método de trabajo es diferente con relación a la experiencia de la persona: *"En efecto, el 60% de los operarios (en realidad los más experimentados) declararon usar la descripción del producto como la clave informativa para iniciar el proceso, mientras que el 40% (los principiantes) confirmaron usar la ubicación del producto para el mismo propósito."* (Caro Gutiérrez, 2018). Esto, también está influenciado por el proceso de capacitación que reciba la persona para cumplir su trabajo en el centro de distribución. (Demaría, 2003)

En resumen, las variables a incluir por este elemento son:

- Usuario
- Antigüedad operario
- Proceso de capacitación (Xlab)
- Hora del alistamiento

#### *Características del producto*

Cada uno de los productos tiene características visuales que lo diferencian de los demás. Adicionalmente, el tiempo que lleve el producto parametrizado en el almacén, corresponde al tiempo en que los operarios están familiarizados con el producto.

*"Existen productos cuyas características de peso, forma o manipuleo dificultan la normal operación del preparador y exige analizar su localización más conveniente, según criterios de ingeniería industrial, ergonometría y seguridad e higiene. En este sentido se corre el riesgo de caídas accidentales de productos sobre otros lo cual, en ocasiones, por las condiciones y exigencias del proceso, no pueden corregirse a tiempo."*

En resumen, las variables a incluir por este elemento son:

- Material
- Antigüedad del material
- Negocio

#### *Características del alistamiento*

El nivel de dificultad de un alistamiento está relacionado con la cantidad de unidades que se deben alistar, y su relación con las características de manipulación de este (Peso y volumen). Cuando la cantidad de unidades y el volumen para manipular sea mayor, aumenta la probabilidad de cometer un error en el conteo de las unidades alistadas. (Caro Gutiérrez, 2018)

En resumen, las variables a incluir por este elemento son:

- Área de picking
- Tipo de almacenamiento
- Nivel de almacenamiento
- Cantidad solicitada
- Peso alistado
- Volumen alistado
- $\frac{1}{4}$  o  $\frac{1}{2}$  caja completa (Un Riesgo)

Finalmente, para completar el dataset a usar, se agrega una variable binaria que identifique cuales alistamientos presentaron novedad en el pasado.

#### *Construcción y exploración del dataset*

El proceso realizado para poder cruzar las tres bases de datos e identificar las variables identificadas inicialmente y mencionadas anteriormente, inicia con la mezcla de las tres bases de datos, el filtrado de los valores que son de interés para el estudio y la depuración de valores faltantes y atípicos. Los pasos realizados fueron los siguientes:

1. Carga de las bases de datos individuales.
2. Filtrado de la base de datos por valores de interés:
  - a. TAP 401: Se dejaron los datos que correspondían al alistamiento de unidades sueltas.
  - b. Unid.: Quitar las unidades de empaque correspondiente a las PQ, que son unidades más pequeñas y que no presentan novedad.
  - c. Serial: Se eliminaron aquellos registros que no tuvieran serial, ya que sin esté no se puede identificar en que contenedor iba y si tenía o no novedad.
3. Mezcla de las bases de datos:
 

Se utiliza merge left sobre la base de datos de alistamientos, para identificar qué alistamiento presentó novedad. La *primary key* corresponde a una clave entre el serial del contenedor y el material alistado.
4. Binarizar *NOVEDADES*:

Independientemente del tipo de novedad, el proyecto está orientado a identificarla. De esta forma se categorizó la variable asignando 1 a aquello que presentó novedad y 0 para los registros contrarios.

**5. Selección de canales de interés:**

Se excluyen canales que no corresponden a la operación actual, que representan volúmenes pequeños o que están “certificados” y no se tienen novedades históricas.

**6. Selección de variables preliminares**

Teniendo en cuenta que la mayoría de la información corresponde a identificadores y guiándonos por las categorías de las causas influyentes, se dejaron las variables útiles para el proyecto:

```
["Material", "Descripción", "ÁreaPicking", "Ubicación", "Centro",  
"Cant.Separada", "PesoAlistado", "VolumenAlistado",  
"Fecha", "Hora", "Mes",  
"Usuario", "Fecha Ingreso", "Xlab",  
"NOVEDAD"]
```

**7. Ajustar formatos de las variables:**

Se realizó el ajuste del tipo de algunas variables (principalmente de fechas) para que fueran reconocidas de forma correcta por el software.

**8. Ajustar nombres de negocio por descripciones:**

En el caso del negocio, este viene codificado así que se realiza el cambio por su descripción para facilitar su entendimiento.

**9. Eliminar datos atípicos:**

Se decide realizar una depuración de valores atípicos identificados. Entre esos están:

- a. Sin unidades alistadas: No se realizó ningún alistamiento.
- b. Con negocio CO19: Este negoció corresponde a material POP y otros insumos que no corresponden a la venta y no se registran novedades.
- c. Sin peso alistado: Hay productos de la misma categoría del CO19 que están categorizados bajo negocios de interés, sin embargo, estos corresponden a insumos similares sin parámetro de venta y peso como los álbumes Jet.

**10. Ajustar datos atípicos y Na's finales**

a. Usuarios Administrativos:

En el caso de los usuarios, todos aquellos que no tuvieran fecha de ingreso corresponden a usuarios administrativos que por algún motivo han realizado confirmaciones del alistamiento de la mercancía. En este caso la fecha de ingreso se colocó como 2018/01/01 y colocando la capacitación (Xlab) como ceros.

b. Ajustar volúmenes en cero:

En algunos casos algunos registros que correspondían a productos comerciales y con peso y cantidades reportadas, no tenían valor del volumen del alistamiento. Se completo esta información identificando el volumen promedio de una unidad con base en la muestra y se estimó el volumen del alistamiento multiplicando este valor por el número de unidades solicitadas.

#### 11. Generación de variables derivadas

La generación de variables aleatorias se genera con el fin de poder evaluar aspectos adicionales que podrían tener influencia en el modelo.

- a. Antigüedad del material  
Se incluyó pensando que un material nuevo puede presentar novedades en el alistamiento debido a que no se conoce físicamente, generando confusión en la cantidad a alistar si no se conoce su empaquetamiento.
- b. Antigüedad de usuarios  
Poder comparar la antigüedad de la persona al momento de hacer el picking.
- c. Hora de trabajo categórica  
En algunas horas del día se presentan cambios de turno, hora de almuerzo, entre otras, que podrían generar novedades en el picking.
- d. Volumen y peso unitario  
Se calculo por aparte el volumen y peso de una unidad del producto, para poder hacer relaciones con otras variables.
- e. Nivel de la ubicación  
Las estructuras del Cartón Flow y Shelving tienen 4 posiciones de forma vertical, donde las posiciones más incómodas para hacer la extracción es la posición superior e inferior.
- f. Media o cuarto de caja  
En el análisis exploratorio se identificó que el 28% de los alistamientos que se realizan por media o un cuarto de caja presenta novedad. El proceso del operario es empezar a extraer unidades para disminuir la manipulación del producto, método que según este dato puede aumentar la probabilidad de presentar novedad. Se calculo como una binaria.

Finalmente, se eliminan varias variables que se habían incluido para poder generar las derivadas y otras porque no se evidencio un impacto con el error en el picking, o están correlacionadas con otras.

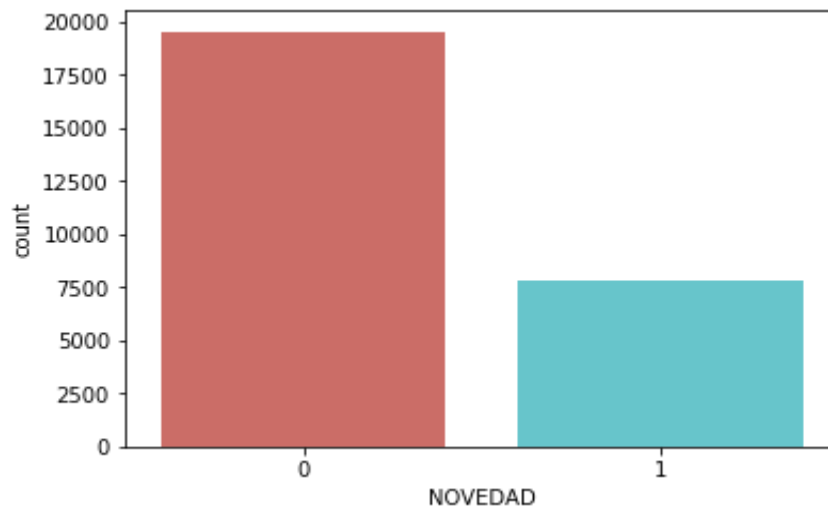
- Peso alistado
- Volumen alistado
- Antigüedad usuario

- Antigüedad material
- Xlab (Capacitación)
- Nivel

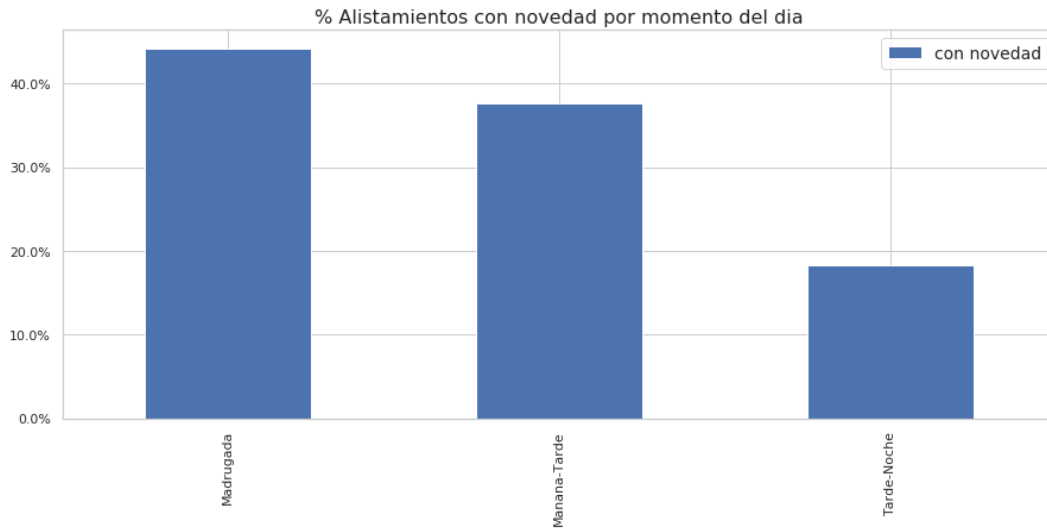
## Reporte de descripción del dataset

Teniendo en cuenta el proceso de construcción, limpieza y exploración del dataset, la base de datos a trabajar en el modelo de clasificación tiene las siguientes características:

- Los atributos numéricos fueron transformados a una escala común, para garantizar que el cálculo de distancias entre registros mantendrá magnitudes “similares” siempre. De esta forma, los modelos de clasificación podrán seleccionar los atributos evitando el sesgo.
- Dada la gran cantidad de registros iniciales (alistanientos) y la ocurrencia de novedades, se apreció un desbalanceo en la variable de clasificación “NOVEDAD”. Para ajustar esta situación, se tomó una muestra más pequeña del set de datos inicial, manteniendo una proporción similar a la inicial, pero que a la vez permitiera encontrar patrones de datos a los algoritmos de clasificación. El resultado se puede apreciar en la ilustración siguiente:



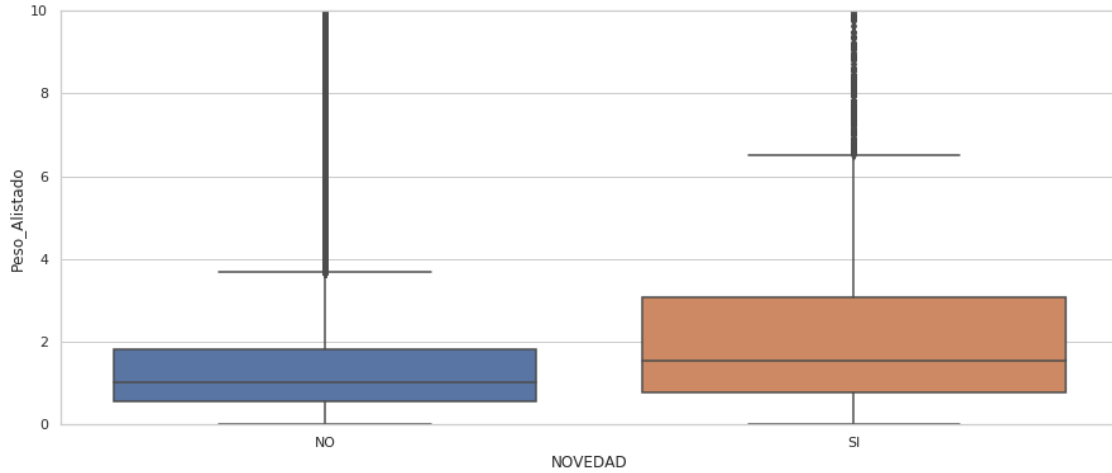
- Como se puede observar en la ilustración X, desde el comienzo del día hasta aproximadamente la 3 de la tarde, ocurren casi el 80 % de las novedades en los alistamientos.



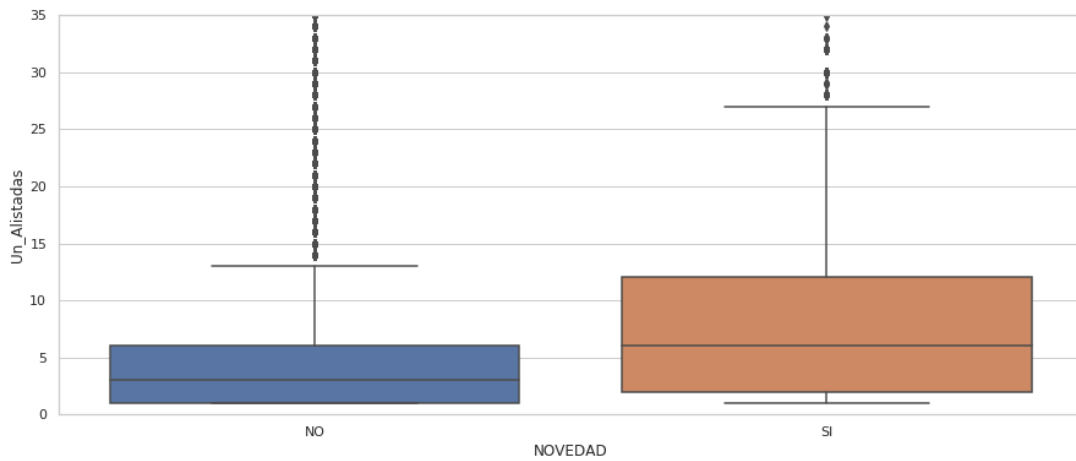
- Los alistamientos para el negocio de chocolate, galletas y café acumulan alrededor del 80% de las incidencias encontradas en el dataset. (Ver ilustración X)



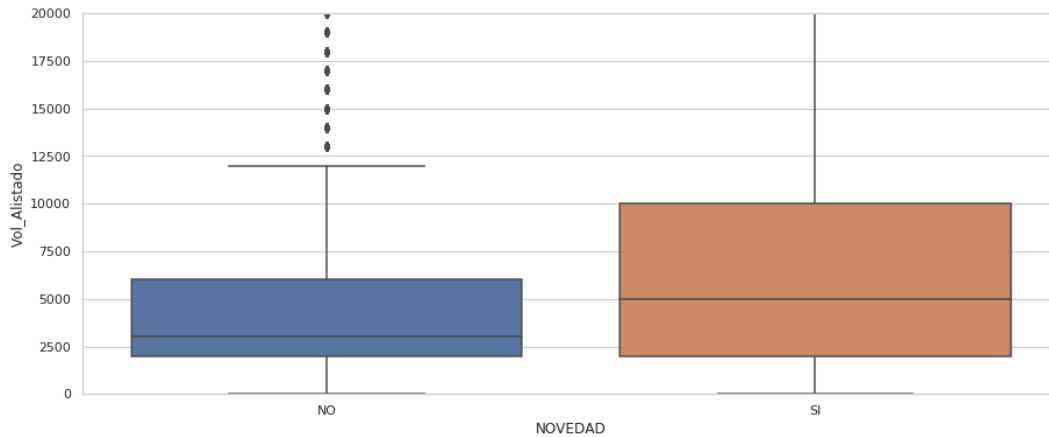
- En promedio, los alistamientos que presentan novedades suelen ser de mayor peso. (Ver ilustración X)



- En promedio, los alistamientos que presentan novedades suelen tener más unidades para alistar. (Ver ilustración X)



- En promedio, los alistamientos que presentan novedades suelen tener un volumen mayor. (Ver ilustración X)



Para iniciar a ejecutar los modelos, se realizó la preparación final evitando repetir actividades costosas en términos de capacidad en cómputo y memoria. El proceso previo a ejecutar el modelo fue el siguiente:

1. Se separaron los predictores de la variable objetivo. De esta forma se obtuvo una matriz de predictores y un vector que indica si hubo novedad en el alistamiento o no.
  2. Se balancearon los datos eliminando registros con valores para las variables predictoras similares. La idea es agrupar en pocos registros la mayor cantidad de información que ofrecen los datos.
  3. Normalizar variables: todas las variables numéricas fueron normalizadas para evitar que los algoritmos de clasificación caigan en sobre ajuste, y puedan ponderar cada atributo de la forma más justa.
  4. Las variables categóricas fueron convertidas a vectores, en donde cada categoría del atributo generó una columna (binaria) en el dataset final. La intención fue crear columnas donde se expresa si una categoría está presente o no en el alistamiento. Esto facilita la aplicación de los modelos de clasificación seleccionados, pues el dataset final solo contiene variables numéricas.
  5. Crear datos de entrenamiento y prueba: se separaron las muestras de entrenamiento y prueba en una proporción de 80/20.
  6. Ejecución de modelos.
- Como consecuencia de lo anterior, el dataset utilizado para ejecutar los modelos cuenta con:
    - 27000 registros.
    - 213 variables.
    - 1 variable o vector de respuesta.



# Modelos

## Diseño del proceso

Para cada modelo propuesto se realizaron varias corridas con miras a encontrar la mejor combinación de hiperparámetros que este alineada con los objetivos de minería. En este caso, se usó la exhaustividad –identificando las novedades- de los modelos como métrica principal de evaluación. Así mismo, se ejecutaron los modelos sobre los mismos sets de entrenamiento y prueba, con el fin de facilitar la comparación de estos.

## Descripción de modelos

### Clasificación de alistamientos

Dado que el set de datos cuenta con muchos predictores, se seleccionaron aquellos modelos que suelen tener buen desempeño frente a bases de datos con dichas características. En ese sentido, se lograron ejecutar los siguientes modelos:

#### Modelo Support Vector Machine:

- Se utilizó un kernel lineal, con costo igual 1 y máximo de iteraciones igual a 4000. Este modelo logró una precisión alta para los alistamientos sin novedad, pero baja para identificar los casos donde sí hay incidencias.

#### Modelo Gradient Boosting:

- Se utilizó una tasa de aprendizaje igual a 0.1, con 100 estimadores y máxima profundidad de árboles igual a 3. Este modelo logró una precisión alta para los alistamientos sin novedad, pero apenas logra identificar los casos donde sí hay incidencias.

#### Modelo Neural Network

- Se implementó una red con 4 capas. Este modelo logró una precisión muy alta para los alistamientos sin novedad y aceptable para identificar los casos donde sí hay incidencias.

## Evaluación y métricas de modelos

#### Modelo Support Vector Machine:

- Métricas sobre el set de entrenamiento

Clase	Precisión	Exhaustividad
Sin Novedad	0.71	0.65
Con Novedad	0.41	0.47

- Métricas sobre el set de prueba

Clase	Precisión	Exhaustividad
Sin Novedad	0.71	0.65
Con Novedad	0.39	0.45

#### Modelo Gradient Boosting:

- Métricas sobre el set de entrenamiento

Clase	Precisión	Exhaustividad
Sin Novedad	0.75	0.96
Con Novedad	0.68	0.34

- Métricas sobre el set de prueba

Clase	Precisión	Exhaustividad
Sin Novedad	0.71	0.95
Con Novedad	0.64	0.30

#### Modelo Neural Network

- Métricas sobre el set de entrenamiento

Clase	Precisión	Exhaustividad
Sin Novedad	0.99	0.99
Con Novedad	0.99	0.99

- Métricas sobre el set de prueba

Clase	Precisión	Exhaustividad
Sin Novedad	0.79	0.77
Con Novedad	0.55	0.59

En general, los modelos de deep learning tuvieron mejor desempeño en términos de los objetivos de minería planteados. Las redes neuronales lograron detectar relaciones e interacciones entre las variables predictoras que no son fáciles de obtener para los modelos de machine learning. A pesar de que los modelos no presentaron una exhaustividad muy alta en ejecuciones sobre sets de

prueba, se decidió que la mejor opción es “Neural Network”. Al aplicarlo, se estaría captando casi el 60 % de las novedades en almacén, con una significativa reducción de los items a revisar.

## Evaluación

Teniendo en cuenta que los objetivos del proyecto están orientados a disminuir el número de novedades desde el picking y mejorar el proceso de verificación de este, se resumirán los resultados en estos aspectos:

### Recomendaciones para el picking

Las recomendaciones del proceso de picking vienen del análisis exploratorio de los datos:

- Trabajar pasando los contenedores en cambios de turno o paradas de descanso.
- Realizar máquina para preembarque para  $\frac{1}{2}$  y  $\frac{1}{4}$  de cajas completas.
  - Actualmente Nutresa cuenta con un equipo de personal reubicado por enfermedad médica dedicada a labores de reempaque y maquila. Este proceso podría ser soportado por este personal y solo debería incurrirse en el gasto del empaque para realizar la maquila.
- Realizar seguimiento de novedades a nivel de detalle del personal del picking.
- Realizar capacitación sobre los materiales con mayor probabilidad de error.

### Mejora de la verificación

Para poder hacer el proceso más eficiente, es importante resumir varios aspectos actuales de este proceso:

- Actualmente hay 7 verificadores por turno (x3) para verificar el 100% de los contenedores con el 100% de los productos que lleva
- Cada verificador le cuesta a Nutresa alrededor de 1'900.000 mes, lo que corresponde alrededor de 480 millones de pesos al año.
- A pesar de revisar el 100% del contenido de todos los contenedores, el 60% presenta novedades. Esto significa que el proceso es efectivo por debajo del 40% dado que se desconoce lo que se envía de más y no se está reportando.
- Los ajustes de inventario mensual cuestan cerca de seis millones.
- En un día promedio se arman cerca de 2500 contenedores, lo que significa que cada operario verificar cerca de 120 contenedores al día.
- Cada contenedor lleva en promedio 7 productos (sku) del cual solicitan en promedio 4.5 unidades, como resultado cada contenedor lleva en promedio 32 unidades y un operario tiene la capacidad de verificar 840 alistamientos en su jornada de trabajo.
- Hay una persona encargada del sistema de cámaras para validar si el error corresponde a un problema del almacén o si es una perdida en calle. Esta cuesta en promedio 2'200.000 al mes.
- El sistema de cámaras cuesta 1'200.000 mensuales por la renovación de 3 de éstas cada mes.

## Indicadores de evaluación

Teniendo en cuenta que se incurren en varios gastos para mantener controlada la pérdida de inventario, se definen los siguientes indicadores para comparar para las propuestas:

- Gasto por contenedor: Corresponde a la distribución de todos los costos identificados sobre el total de contenedores alistados.
- % Efectividad: Corresponde a los contenedores identificados con novedad del total que presentan novedad.
- ROI: Retorno de la inversión por el ahorro de la propuesta.

## Modelo de Machine / Deep Learning

La primera opción es poder aplicar una red neuronal con el fin de realizar una verificación selectiva de los contenedores, evitando realizar una revisión del 100% de los contenedores y realizando verificaciones puntuales de algunos de los alistamientos que están en el contenedor.

Las premisas y supuestos de esta opción son las siguientes:

- El modelo requiere revisar cerca del 23% de todos los alistamientos para encontrar cerca del 60% de los alistamientos con novedad.
- Se requiere definir una nueva forma de captura de información para las novedades reportadas en los muelles y en la calle, ya que la información es manual.
- El modelo debe tener la capacidad de actualizarse debido a la rotación de personal y a la creación de nuevos materiales a lo largo del año.
- Es necesario programar una herramienta. Esta puede estar instalada en un computador local pero debe cumplir con las condiciones mencionadas anteriormente.
- El modelo trabajaría con la información de los últimos tres meses de operación para ir actualizándose frente a cualquier tendencia de personal, productos, ubicación, slotting, entre otros.
- El proceso de verificación no se realizaría en los muelles sino al final del sistema de picking, lugar donde es más fácil y rápido realizar ajustes al alistamiento.
- El auditor de cámaras y los elementos de esto siguen utilizándose como evidencia que presente cualquier aliado de transportes.

Para el desarrollo de una herramienta con características a las descritas, SMART DECISION SAS estima que una propuesta de este tipo puede estar costando alrededor de los 30 millones de pesos para programar la herramienta sin conectividad al WMS y al WCS de Nutresa, pero permitiendo al analista de slotting o inventarios manipularla.

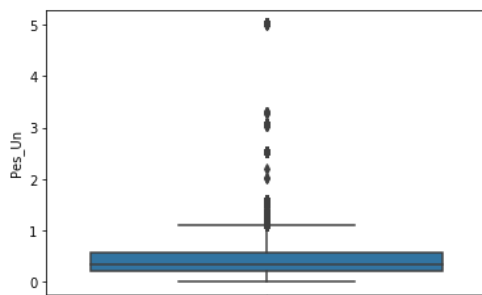
## Máquina de control de peso

El sistema de control de peso es un mecanismo mecánico que toma el peso del contenedor y de acuerdo con los materiales y cantidades que lleva, realiza una verificación del peso real contra el

peso teórico para saber si existen diferencias. En caso de encontrarlas, desvía el contenedor para su revisión y ajuste.

Las premisas y supuestos de esta opción son:

- Se supone que aproximadamente el 1% de los contenedores se han identificado con novedades. Por ello se estima que con esta opción el 8% de los contenedores deberán ser verificados teniendo en cuenta que no se tiene data de las novedades no reportadas por sobrantes.
- El sistema debe tomar peso de la caja antes de iniciar la inducción del producto, para realizar un ajuste de peso por diferencia del peso del contenedor por tamaño o material.
- El sistema de la máquina debe estar conectado al WMS para conocer el peso teórico del contenedor y al WCS para desviarlo desde los conveyors del sistema.
- El proceso de verificación no se realizaría en los muelles sino al final del sistema de picking, donde se ubicaría el control de peso que compara el teórico vs el real ajustando el peso del corrugado.
- El auditor de cámaras y los elementos de esto siguen utilizándose como evidencia que presente cualquier aliado de transportes.
- El percentil 25 del peso de una unidad corresponde a los 220 gr. Tener en cuenta esto como tolerancia de error de la máquina.



Para este desarrollo se tiene una cotización de INTECOL, la cual contempla la máquina, la banda transportadora para desviar contenedores con novedades y la programación con el sistema WMS y WCS. Esta cotización está por un costo de USD 65,717.

### Análisis de propuestas

El resumen de las propuestas evidencia una mejora sustancial del costo que asume cada contenedor disminuyéndolo en más de un 62% para el caso del ML y en un 84% para en caso del control de peso.

	Actual	Modelo ML	Control de peso
<b>PARAMETROS</b>			
Contenedores mes	60,000	60,000	60,000
% Verificación	100%	30%	8%
Alistamientos verificados	420,000	126,000	33,600
Operarios Requeridos	21.00	7.00	2.00

<b>GASTOS MENSUALES</b>			
Operarios	\$ 39,900,000	\$ 13,300,000	\$ 3,800,000
Auditor cámaras	\$ 2,200,000	\$ 2,200,000	\$ 2,200,000
Renovación cámaras	\$ 1,200,000	\$ 1,200,000	\$ 1,200,000
<b>Costo Total</b>	<b>43,300,000</b>	<b>16,700,000</b>	<b>7,200,000</b>

<b>INDICADORES</b>			
<b>Costo x Cont</b>	\$ 721.67	\$ 278.33	\$ 120.00
<b>% Efectividad</b>	40%	60%	100%

<b>AHORRO Y ROI</b>			
<b>Inversión</b>	\$ -	\$ 30,000,000	\$ 221,466,290
<b>Ahorro Anual</b>	-	319,200,000	433,200,000
<b>ROI (meses)</b>	-	1.13	6.13

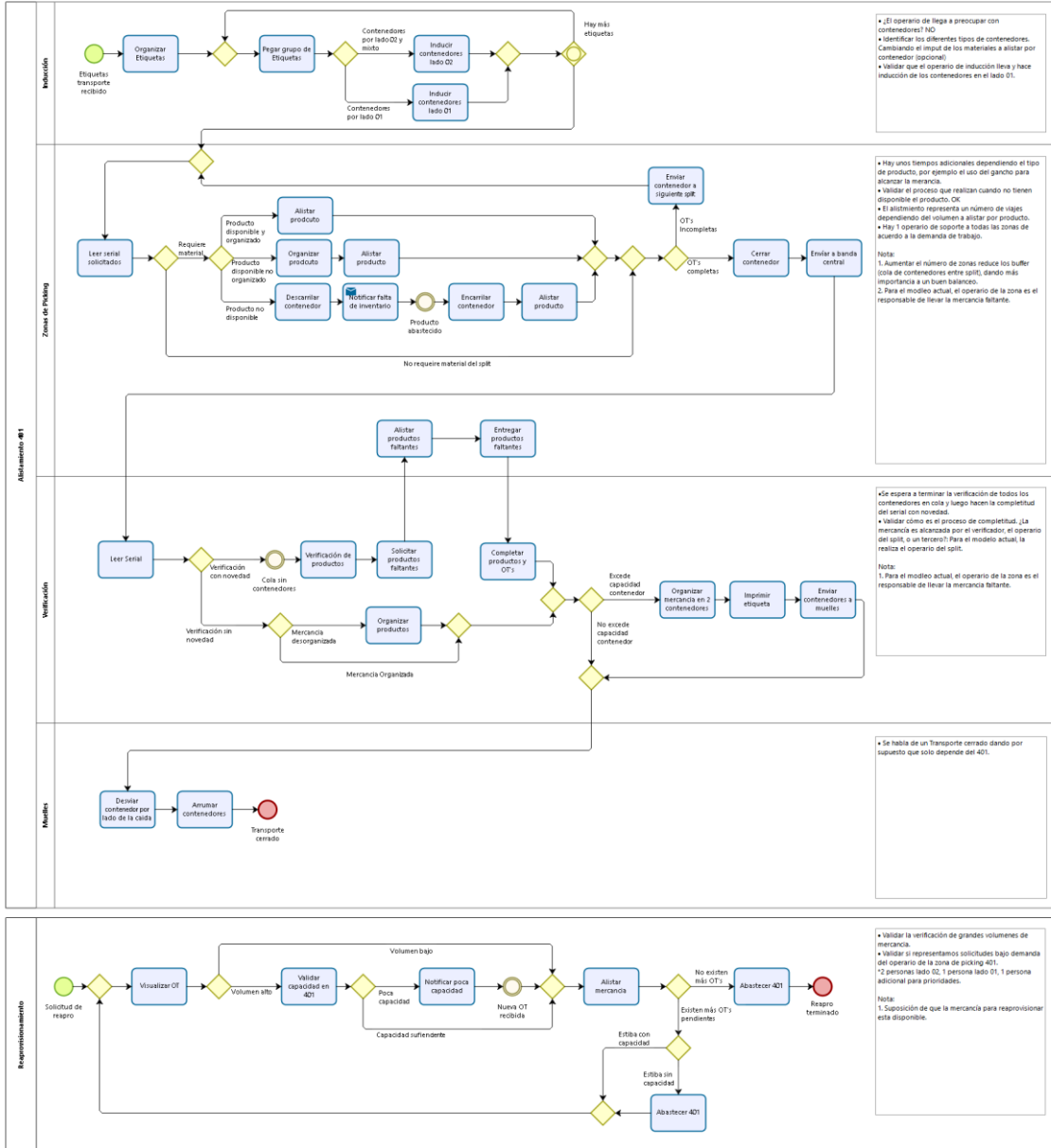
Cada una de las alternativas tiene diferentes bondades frente al proceso de verificación y frente al nivel de la inversión y el retorno de esta, sin embargo, ambas presentan resultados favorables para mejorar el proceso de verificación mejorando el costo por servir.

## Anexos

### Anexo 1. Glosario de Terminología

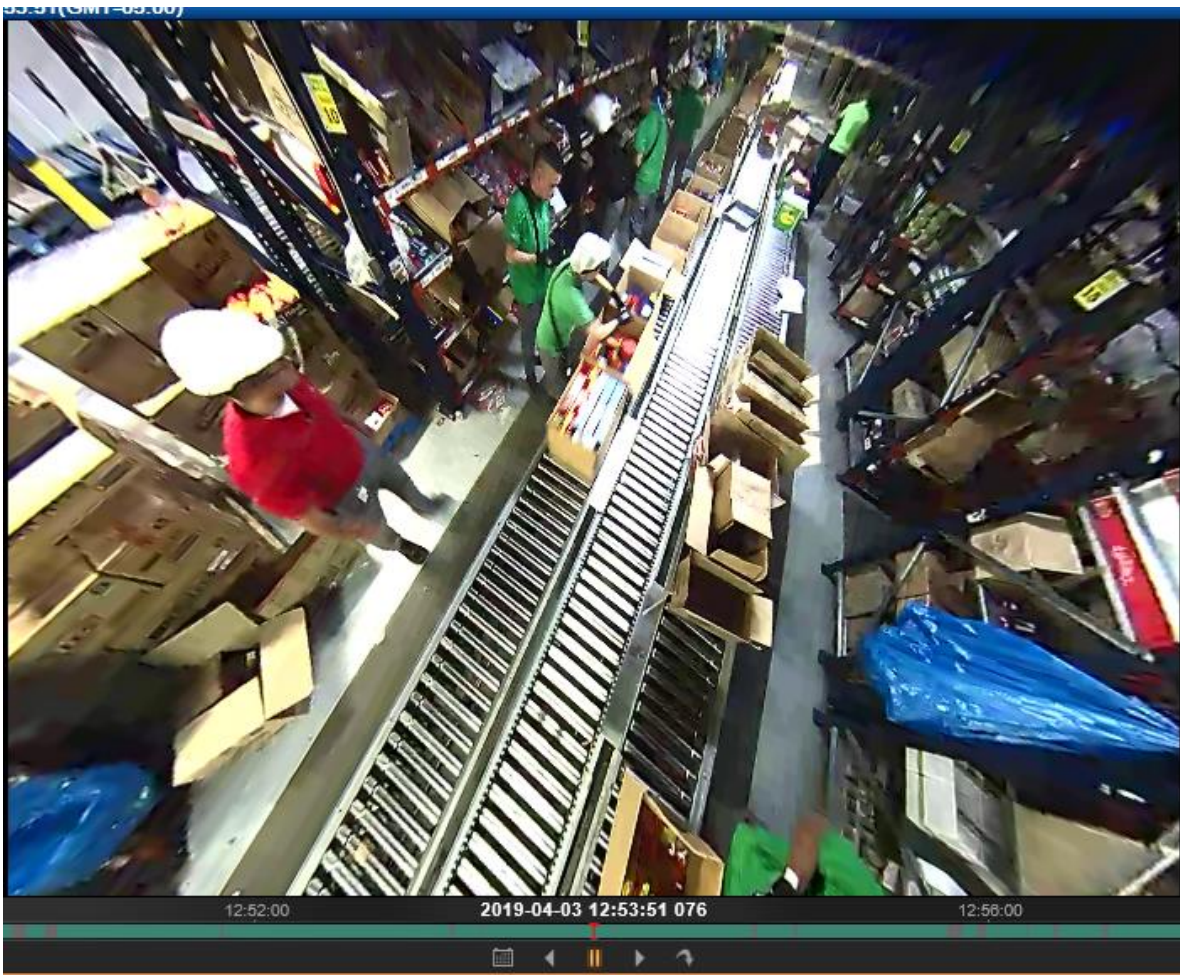
- ÁREA DE PICKING=Espacio de trabajo para un operario.
- CEDI=Centro de distribución.
- DESPACHO=Cargue y distribución de la mercancía en algún medio de transporte.
- ENSAMBLE PROGRESIVO=Contenido de un pedido que pasa de una zona de alistamiento tras otra hasta completar el pedido.
- FRECUENCIA=número de alistamientos del material.
- GAVETA=Recipiente o caja en el cual se depositan las unidades de materiales pedidas.
- INVENTARIO DE PICKING=Ubicaciones de almacenamiento de las cuales se realiza el proceso de o alistamiento.
- LINEAS DE PEDIDO=Número de referencias incluidas en un mismo pedido.
- MATERIAL=Producto identificado con un SKU.
- OPERARIO=Recurso humano que mueve los materiales en un CEDI.
- ORDEN DE TRABAJO=Solicitud de alistamiento de cierta cantidad de unidades por material.
- PERFIL=Conjunto de rasgos de las actividades del almacén.
- PICKING=Alistamiento de los materiales de un pedido.
- PICK TO BETL=Alistamiento de mercancía donde el operario hace la inducción de las unidades en una caja la cual es transportada sobre una banda transportadora hacia los muelles de cargue.
- PROFUNDIDAD DE LÍNEA=Valor esperado de unidades a alistar cada vez que se pide un material.
- SISTEMA MECANIZADO=Estructura mecánica diseñada para el alistamiento de pedidos en forma de ensamble progresivo.
- SKU=Stock keeping unit.
- SLOTTING=Ubicación estratégica de materiales.
- STORAGE=Acomodación de la mercancía en las ubicaciones.
- UBICACIÓN=Espacio físico para la ubicación de materiales.
- UNIDAD DE MANIPULACIÓN=Cajas, estibas, bultos, unidades.
- VOLUMEN= Metros cúbicos pedidos de un material.

## Anexo 2. Diagrama de procesos de picking





Anexo 3. Sistema de vídeo para verificación de errores en picking



## Bibliografía

- Anaya Tejero, J. J. (2008). *Almacenes. Análisis, diseño y organización*. ESIC.
- Caro Gutiérrez, M. P. (febrero de 2018). El error humano en la operación de picking. Una aproximación metodológica desde la ergonomía cognitiva. Bogotá, Colombia.
- Demaría, G. J. (06 de 2003). *Errores en el Picking*. Obtenido de <http://www.logisticamx.enfasis.com/notas/3753-errores-el-picking>
- Gómez, J. V. (s.f.). *Expertos en marca*. Obtenido de La magia de los agotados en el punto de venta: <https://expertosenmarca.com/la-magia-de-los-agotados-en-el-punto-de-venta/?amp>
- Logística, Z. (6 de 11 de 2018). *Entregas certificadas un proposito fundamental*. Obtenido de <https://www.zonalogistica.com/entregas-certificadas-un-proposito-fundamental-del-ecr/>
- Logyca. (26 de 05 de 2017). Obtenido de ¿Qué es la venta perdida?: <http://blog.logyca.com/nosotros/que-es-venta-perdida/>
- Mojica, D. (03 de 2019). Jefe de Almacén y Transporte. (S. Tovar, Entrevistador)
- Nutresa, G. (2017). *Presentación corporativa Grupo Nutresa 2017*. Obtenido de <https://www.gruponutresa.com/wp-content/uploads/2015/12/Presentaci%C3%B3n-corporativa-Grupo-Nutresa-2017.pdf>
- Nutresa, G. (2019). *Quiénes somos - Grupo Nutresa*. Obtenido de <https://www.gruponutresa.com/quienes-somos/>