



# Pronóstico de la demanda para una cadena de tiendas de ropa y calzado

**Proyecto final de máster**

ARTIFICIAL INTELLIGENCE

THREE POINTS. THE DIGITAL BUSINESS SCHOOL

**Grupo 2**

Christian Arturo Bascompte Llerena

Juan Carlos Jiménez Bedolla

Luis Alejandro Maguey Del Razo

Luis Fernando Quesada Herra

*June 22, 2020*

THREE POINTS

La propiedad intelectual pertenece a Price Shoes, se prohíbe su reproducción en cualquier medio físico o electrónico su no es con consentimiento expreso de Price Shoes.

*Abril, 2020*

## Resumen ejecutivo

Three Points The Digital Business School, ubicada en la ciudad de Barcelona, España, y que es parte del Grupo Planeta, Formación y Universidades, es una escuela de negocios digitales que da respuesta a las necesidades del escenario económico y empresarial actual, que tiene como objetivo fundamental el asegurar la competitividad de sus estudiantes y apoyar a su proyección profesional por medio de una propuesta académica que está alineada con las necesidades de las empresas adaptando los perfiles a las nuevas formas de trabajar (que son más participativas y eficientes) y enfocados en las exigentes demandas del consumidor digital.

Dentro de los programas que ofrece, se encuentra el executive en inteligencia artificial el cual se enfoca en la aplicación de tecnologías como lo son el aprendizaje automático o el aprendizaje profundo, las redes neuronales y el uso de framework en distintas plataformas las cuales se pueden aplicar de manera viable a problemas de clasificación y pronóstico.

Surge entonces la iniciativa de aplicar dicha tecnología en un modelo de negocio dedicado a venta de artículos de moda de ropa y calzado, particularmente en el pronóstico de la demanda y clasificación de artículos según sus características físicas.

Esto implica, un cambio de paradigma en la aplicación de los métodos actuales que tienen los planeadores de la demanda donde hoy se consideran ciertos criterios que afectan al comportamiento de las ventas de los diferentes artículos como son: temporadas, características físicas, tendencias en la moda, clima, presentación en los catálogos, número de catálogos en los que aparece, cantidad de catálogos vendidos, variedad de colores, modelos similares, cantidad de artículos dentro de la misma familia, clasificación del artículo de acuerdo a modelos anteriores, etc.

Es de esperarse que toda esa información es difícil de clasificar y verificar qué peso tiene en cada modelo dentro de 8.000 variedades que se venden en las 17 bodegas de la cadena y en los 50 catálogos que se editan a lo largo del año.

Como parte del programa de formación del programa, una de las propuestas de valor se concentra en la aplicación práctica de los conocimientos adquiridos por medio de un proyecto final. La propuesta de aplicar la tecnología de la inteligencia artificial en una cadena de tiendas de moda fue propuesta al comité evaluador y fue aceptada como proyecto final del máster.

La cadena de tiendas de moda es Price Shoes que está ubicada en México y tiene una presencia en todo el país consolidándose como una empresa con un historial de más de 20 años en el mercado con más de 50 tiendas en diferentes formatos, dos centros de distribución y más de 700.000 asociados. Es una de las principales cadenas de calzado y ropa de latinoamérica y líder en venta por catálogo.



# Índice general

<b>1</b>	<b>Introducción</b>	<b>7</b>
1.1	Motivación	7
1.2	Objetivos	8
1.3	Contexto	8
1.3.1	Referencias	9
<b>2</b>	<b>Metodología</b>	<b>11</b>
<b>2.1</b>	<b>CRISP-DM</b>	<b>11</b>
2.1.1	Fase I. Business Understanding (Comprensión del negocio)	12
2.1.2	Fase II. Data Understanding (Estudio y comprensión de los datos)	12
2.1.3	Fase III. Data Preparation (Análisis de los datos y selección de características)	13
2.1.4	Fase IV. Modeling (Modelado)	13
2.1.5	Fase V. Evaluation (Obtención y evaluación de resultados)	13
2.1.6	Fase VI. Deployment (Despliegue o puesta en producción)	13
<b>2.2</b>	<b>Agile</b>	<b>13</b>
<b>2.3</b>	<b>Kanban</b>	<b>14</b>
2.3.1	Definir el flujo de trabajo de los proyectos	15
2.3.2	Visualizar la fases del ciclo de producción	15
2.3.3	Stop starting, starting finishing	16
2.3.4	Control del flujo	16
<b>2.4</b>	<b>Github</b>	<b>16</b>
<b>3</b>	<b>Planificación del trabajo</b>	<b>19</b>

---

<b>4</b>	<b>Selección de los modelos</b>	<b>21</b>
<b>4.1</b>	<b>Datos</b>	<b>21</b>
4.1.1	Generar Datasets	21
4.1.2	Carga inicial de datos	22
4.1.3	Limpieza de Datos	23
<b>4.2</b>	<b>Pronóstico de la demanda</b>	<b>23</b>
<b>5</b>	<b>Construcción y entrenamiento de los modelos</b>	<b>27</b>
<b>6</b>	<b>Métricas y evaluación de los modelos</b>	<b>29</b>
<b>7</b>	<b>Resultados</b>	<b>31</b>
<b>8</b>	<b>Discusión</b>	<b>33</b>
<b>9</b>	<b>Conclusiones</b>	<b>35</b>





# 1. Introducción

## 1.1 Motivación

En la junta de estrategias anuales de Dirección General de Price Shoes en 2018, el **Sr. José Schatz** presidente ejecutivo de la empresa remarcó:

La empresa debe apoyarse en la tecnología para aprovecharla en la automatización de las actividades manuales y generar un diferenciador con nuestra competencia y ofrecer más y mejores servicios a nuestros clientes. Quiero tener tecnología de punta e invertiré lo necesario para lograrlo.

Así que el departamento de desarrollo de sistemas debe afrontar el reto y proponer nuevas herramientas que permitan cumplir las expectativas de la Dirección General y alinearse a la visión corporativa.

El pronóstico de la demanda es la piedra angular de las compras y el resurtido de la mercancía que se vende en las tiendas, cualquier desviación se traduce en pérdidas importantes por el desabasto o sobre inventario.

El compromiso de Prices Shoes ante sus socios es tener disponible la mercancía de todos los catálogos en todo el tiempo de vigencia.

La misión de Price Shoes es

Servir a la sociedad, poniendo al alcance de la gente un negocio propio, a través del exclusivo sistema de ventas por catálogo de Price Shoes; el beneficio para las personas involucradas es el de poder convertirse en microempresarios exitosos, a los que nadie detenga su crecimiento, considerando que la empresa los apoyará para que obtengan beneficios económicos, personales y profesionales.

Por lo tanto, se vuelve muy importante tener inventarios sanos en las tiendas y una de las dependencias más importantes es la correcta planeación de la demanda además del suministro.

## 1.2 Objetivos

Los objetivos generales del presente trabajo están orientados a resolver las necesidades de la demanda que actualmente se hace de manera manual con herramientas no eficientes y que son muy susceptibles de errores.

- Reducir la desviación entre la demanda y la venta por artículo
- Aumentar la productividad del área de demanda ajustando los artículos en el momento que se detecta la desviación.
- Desarrollar soluciones de Machine Learning que sean capaces de clasificar los artículos nuevos de acuerdo a sus características físicas para ubicarlos en la segmentación de venta correcta.
- Desarrollar solución que sea capaz de hacer el pronóstico de la demanda para artículos nuevos y de resurtido.

Los objetivos específicos están orientados a tener predicciones más precisas y que consideren las variables que impactan directamente al comportamiento de ventas de los artículos, conocer la magnitud del impacto de cada variable y que se apliquen los ajustes necesarios para tener una precisión que no se degrade con el aumento de las tiendas o artículos.

- Proponer modelos de Machine Learning para determinar las variables más importantes que influyen en la venta de los artículos.
- Proponer modelos que mediante un entrenamiento puedan determinar los pesos específicos que tiene cada variable en el pronóstico particular de cada artículo.
- Proponer modelos que determinen la clasificación de un modelo como Bueno, Regular o Riesgo analizando las características de la imagen del artículos
- Proponer modelos de Machine Learning para ajustar las demandas de los artículos de acuerdo al comportamiento de ventas.

## 1.3 Contexto

Price Shoes es una cadena de bodegas más grande del mundo de calzado y ropa, está consolidada en la venta por catálogo y tiene una gran cantidad de asociados que usan el camabaceo como medio de promoción, venta y cobro.

La socia adquiere un conjunto de catálogos y adquiere la membresía que le da precios preferenciales de mayoreo para que ofrezca a sus clientes precios competitivos, se capacita a las socias para que la venta sea más efectiva y pueda tener un negocio propio como medio de ingresos personales.

Cada catálogo que se vende a las socias en una oportunidad de ofrecer los artículos a sus clientes finales, por lo tanto, entre más catálogos tenga ofrecerá la mercancía a más clientes y tendrá más oportunidades de venta. La ecuación es sencilla, entre más socios se tengan y cada socio tenga más catálogos se tendrá más ventas.

El abasto de la mercancía es una de las prioridades principales de Price Shoes, ya que los socios requieren de la existencia en el momento que vayan por sus pedidos. Mientras el



catálogo esté vigente la mercancía debe estar en las tiendas en las cantidades suficientes para que no haya desabasto y tampoco excedentes; al final del catálogo la mercancía debe terminarse también 30 días posteriores a la vigencia final.

Las ventas que se generan en las tiendas son procesadas en los servidores locales y se replican en procesos automáticos nocturnos a la base de datos central, de tal manera que a las 12pm se cuenta con la información de ventas y existencias de todas las tiendas disponible para procesos en el corporativo.

### 1.3.1 Referencias

Price Shoes cuenta con diferentes redes sociales como Facebook, Instagram, Twiter y también realiza ventas en línea en su página de internet.



Página de internet de Price Shoes <https://www.priceshoes.com>





## 2. Metodología

Como tal no existe una Metodología específica para un proyecto de IA sin embargo se toman como base algunos de los principales marcos Metodológicos o *frameworks* que actualmente se utilizan para el desarrollo de aplicaciones de *Data Mining* que tienen una evolución aplicable a las capacidades de IA que se persiguen en el presente proyecto.

Una metodología es útil en un proyecto porque proporciona un marco donde se pueden aplicar lecciones aprendidas de proyectos anteriores asegurando el correcto desarrollo del mismo. Existen metodologías para la gestión de los recursos y fases del proyecto como por ejemplo el Project Managment que divide el proyecto en 5 grupos de procesos y 13 áreas de conocimiento. En conjunto lleva la gestión por cada uno de los 5 procesos de la vida del proyecto y su correcta gestión. El objetivo del presente trabajo se centra en el desarrollo e implementación de la propuesta de solución, así que las metodologías propuestas están orientadas a éstos dos aspectos.

Primero se exponen las metodologías que nos ayudan a entender las necesidades del cliente y luego el desarrollo y validación de la solución. Se propone el uso de **CRISP-DM** para la evolución de las soluciones y para el desarrollo se propone la metodología **Agile** para el desarrollo que permite interactuar al usuario con la solución y permite tener contacto y retro alimentación desde las primeras liberaciones. Para llevar el control de la metodología Agile se usará la metodología **Kanban** la cual es muy sencilla de implementar y seguir dando claridad del avance del proyecto dado que no se puede hacer un cronograma riguroso por medio de algún gestor como Microsoft Project, en su lugar usaremos la plataforma que ofrece **Github** para llevar la documentación desde el diccionario de términos o wiki, repositorio de códigos con control de versiones, hasta el tablero Kanban.

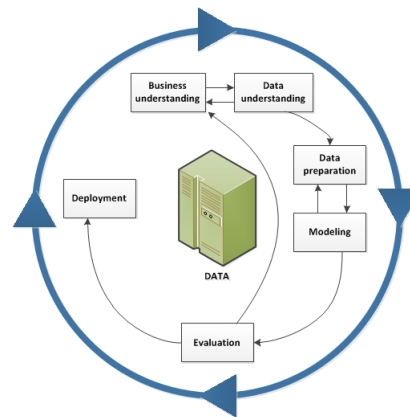
### 2.1 CRISP-DM

Los orígenes de CRISP-DM son en el año 1999 cuando un importante consorcio de empresas europeas tales como NCR (Dinamarca), AG(Alemania), SPSS (Inglaterra), OHRA (Holanda), Te-

radata, SPSS, y Daimler-Chrysler, proponen a partir de diferentes versiones de KDD (Knowledge Discovery in Databases) [Reinartz, 1995], [Adraans, 1996], [Brachman, 1996], [Fayyad, 1996], el desarrollo de una guía de referencia de libre distribución denominada CRISP-DM (Cross Industry Standard Process for Data Mining).

CRISP-DM proporciona una descripción normalizada del ciclo de vida de un proyecto estándar de análisis de datos, de forma análoga a como se hace en la ingeniería del software con los modelos de ciclo de vida de desarrollo de software. El modelo CRISP-DM cubre las fases de un proyecto, sus tareas respectivas, y las relaciones entre éstas.

El ciclo de vida del proyecto de minería de datos consiste en seis fases mostradas en la figura siguiente.



La secuencia de las fases no es rígida, se permite movimiento hacia adelante y hacia atrás entre diferentes fases. El resultado de cada fase determina qué fase, o qué tarea particular de una fase, hay que hacer después. Las flechas indican las dependencias más importantes y frecuentes.

El círculo externo en la figura simboliza la naturaleza cíclica de los proyectos de análisis de datos. El proyecto no se termina una vez que la solución se despliega. La información descubierta durante el proceso y la solución desplegada pueden producir nuevas iteraciones del modelo. Los procesos de análisis subsecuentes se beneficiarán de las experiencias previas.

### 2.1.1 Fase I. Business Understanding (Comprensión del negocio)

Esta primer fase está orientada a la comprensión de las necesidades del negocio. Después se convierte este entendimiento en la definición de los objetivos del proyecto y en un plan preliminar diseñado para lograrlos.

### 2.1.2 Fase II. Data Understanding (Estudio y comprensión de los datos)

La fase de entendimiento de datos comienza con la recolección de datos inicial y continúa con las actividades que permiten familiarizarse con ellos, identificar los problemas de calidad, descubrir

conocimiento preliminar sobre los datos, y/o descubrir subconjuntos interesantes para formar hipótesis en cuanto a la información oculta.

### 2.1.3 Fase III. Data Preparation (Análisis de los datos y selección de características)

La fase de preparación de datos cubre todas las actividades necesarias para construir el conjunto final de datos (los datos que se utilizarán en las herramientas de modelado) a partir de los datos en bruto iniciales. Las tareas incluyen la selección de tablas, registros y atributos, así como la transformación y la limpieza de datos para las herramientas que modelan.

### 2.1.4 Fase IV. Modeling (Modelado)

En esta fase, se seleccionan y aplican las técnicas de modelado que sean pertinentes. Típicamente hay varias técnicas para el mismo tipo de problema de minería de datos. Algunas técnicas tienen requerimientos específicos sobre la forma de los datos. Por lo tanto, casi siempre en cualquier proyecto se acaba volviendo a la fase de preparación de datos.

### 2.1.5 Fase V. Evaluation (Obtención y evaluación de resultados)

En esta etapa en el proyecto, se han construido uno o varios modelos que parecen alcanzar calidad suficiente desde la una perspectiva de análisis de datos.

Antes de proceder al despliegue final del modelo, es importante evaluarlo a fondo y revisar los pasos ejecutados para crearlo, comparar el modelo obtenido con los objetivos de negocio. Un objetivo clave es determinar si hay alguna cuestión importante de negocio que no haya sido considerada suficientemente. Al final de esta fase, se debería obtener una decisión sobre la aplicación de los resultados del proceso de análisis de datos.

### 2.1.6 Fase VI. Deployment (Despliegue o puesta en producción)

Generalmente, la creación del modelo no es el final del proyecto. Incluso si el objetivo del modelo es de aumentar el conocimiento de los datos, el conocimiento obtenido tendrá que organizarse y presentarse para que el cliente pueda usarlo. Dependiendo de los requisitos, la fase de desarrollo puede ser tan simple como la generación de un informe o tan compleja como la realización periódica y quizás automatizada de un proceso de análisis de datos en la organización.

## 2.2 Agile

Las metodologías ágiles han venido a solucionar la reducción del tiempo de entrega y el contacto del usuario final con el producto terminado o semi-terminado lo cual permite una retro-alimentación temprana para hacer los ajustes necesarios a tiempo y no esperar hasta el producto totalmente terminado.

Por definición, las metodologías ágiles son aquellas que permiten adaptar la forma de trabajo a las condiciones del proyecto, consiguiendo flexibilidad e inmediatez en la respuesta para amoldar el



proyecto y su desarrollo a las circunstancias específicas del entorno.

En esencia, las empresas que apuestan por esta metodología consiguen gestionar sus proyectos de forma flexible, autónoma y eficaz reduciendo los costes e incrementando su productividad.

Las metodologías ágiles mejoran la satisfacción del cliente dado que se involucrará y comprometerá a lo largo de todo el proyecto. En cada etapa se informará al cliente de los logros y progresos del mismo, con la visión de involucrarlo directamente para sumar su experiencia y conocimiento, y así, optimizar las características del producto final obteniendo en todo momento una visión completa de su estado.

Otra de las ventajas es la mejora de la motivación e implicación del equipo de desarrollo. Pero esta mejora no es casual: las metodologías ágiles permiten a todos los miembros del equipo conocer el estado del proyecto en cualquier momento, así, los compromisos son negociados y aceptados por todos los miembros del equipo.

Por otro lado, cabe destacar que optar por la aplicación de una gestión ágil permite ahorrar tiempo y costes. El desarrollo ágil trabaja de un modo más eficiente y rápido, y con ello, se cumple de forma estricta el presupuesto y los plazos pactados dentro de un proyecto.

Se trabaja con mayor velocidad y eficiencia. Una de las máximas de su aplicación es que se trabaja a través de entregas parciales del producto, de este modo, es posible entregar en el menor intervalo de tiempo posible una versión mucho más funcional del producto.

Gracias a las entregas parciales (centradas en entregar en primer lugar aquellas funcionalidades que aportan valor) y a la implicación del cliente será posible eliminar cualquier característica innecesaria del producto.

Las metodologías ágiles permiten mejorar la calidad del producto. La continua interacción entre los desarrolladores y los clientes tiene como objetivo asegurar que el producto final sea exactamente lo que el cliente busca y necesita.

## 2.3 Kanban

Actualmente, el término Kanban ha pasado a formar parte de las llamadas metodologías ágiles, cuyo objetivo es gestionar de manera general cómo se van completando las tareas. Kanban es una palabra japonesa que significa “tarjetas visuales”, donde Kan es “visual”, y Ban corresponde a “tarjeta”. Esta consiste en la elaboración de un cuadro o diagrama en el que se reflejan tres columnas de tareas; pendientes, en proceso o terminadas. Este cuadro debe estar al alcance de todos los miembros del equipo, evitando así la repetición de tareas o la posibilidad de que se olvide alguna de ellas. Por tanto, ayuda a mejorar la productividad y eficiencia del equipo de trabajo.

La metodología Kanban se basa en una serie de principios que la diferencian del resto de metodologías ágiles:

- **Calidad garantizada.** Todo lo que se hace debe salir bien a la primera, no hay margen de error. De aquí a que en Kanban no se premie la rapidez, sino la calidad final de las tareas realizadas. Esto se basa en el hecho que muchas veces cuesta más arreglarlo después que hacerlo bien a la primera.
- **Reducción del desperdicio.** Kanban se basa en hacer solamente lo justo y necesario, pero hacerlo bien. Esto supone la reducción de todo aquello que es superficial o secundario (principio YAGNI).
- **Mejora continua.** Kanban no es simplemente un método de gestión, sino también un sistema de mejora en el desarrollo de proyectos, según los objetivos a alcanzar.
- **Flexibilidad.** Lo siguiente a realizar se decide del backlog (o tareas pendientes acumuladas), pudiéndose priorizar aquellas tareas entrantes según las necesidades del momento (capacidad de dar respuesta a tareas imprevistas).

La aplicación del método Kanban implica la generación de un tablero de tareas que permitirá mejorar el flujo de trabajo y alcanzar un ritmo sostenible.

### 2.3.1 Definir el flujo de trabajo de los proyectos

Para ello, simplemente deberemos crear nuestro propio tablero, que deberá ser visible y accesible por parte de todos los miembros del equipo. Cada una de las columnas corresponderá a un estado concreto del flujo de tareas, que nos servirá para saber en qué situación se encuentra cada proyecto. El tablero debe tener tantas columnas como estados por los que pasa una tarea, desde que se inicia hasta que finaliza (por ejemplo: diagnóstico, definición, programación, ejecución, testing, etc.).

A diferencia de SCRUM, una de las peculiaridades del tablero es que este es continuo. Esto significa que no se compone de tarjetas que se van desplazando hasta que la actividad queda realizada por completo. En este caso, a medida que se avanza, las nuevas tareas (mejoras, incidencias o nuevas funcionalidades) se acumulan en la sección inicial, de manera que en las reuniones periódicas con el cliente se priorizan y se colocan dentro de la sección que se estima oportuna.

Dicho tablero puede ser específico para un proyecto en concreto o bien genérico. No hay unas fases del ciclo de producción establecidas sino que se definirán según el caso en cuestión, o se establecerá un modelo aplicable genéricamente para cualquier proyecto de la organización.

### 2.3.2 Visualizar la fases del ciclo de producción

Al igual que Scrum, Kanban se basa en el principio de desarrollo incremental, dividiendo el trabajo en distintas partes. Esto significa que no hablamos de la tarea en sí, sino que lo dividimos en distintos pasos para agilizar el proceso de producción.

Normalmente cada una de esas partes se escribe en un post-it y se pega en el tablero, en la fase que corresponda. Dichos post-its contienen la información básica para que el equipo sepa rápidamente la carga total de trabajo que supone: normalmente descripción de la tarea con la estimación de horas. Además, se pueden emplear fotos para asignar responsables así como también usar tarjetas con distintas formas para poner observaciones o indicar bloqueos (cuando una tarea no puede hacerse porque depende de otra).

Al final, el objetivo de la visualización es clarificar al máximo el trabajo a realizar, las tareas asignadas a cada equipo de trabajo (o departamento), así como también las prioridades y la meta asignada.

### 2.3.3 Stop starting, starting finishing

Este es el lema principal de la metodología Kanban. De esta manera, se prioriza el trabajo que está en curso en vez de empezar nuevas tareas. Precisamente, una de las principales aportaciones del Kanban es que el trabajo en curso debe estar limitado y, por tanto, existe un número máximo de tareas a realizar en cada fase.

En realidad, se trata de definir el máximo número de tareas que podemos tener en cada una de las fases (p.e: 3 tareas en la fase de planificación; 2, en la fase de desarrollo; una, en la fase de pruebas, etc.) y, por tanto, restringir el trabajo en curso. A esto, se le añade otra idea que, por muy obvia que pueda parecer, la práctica nos demuestra que no es así: no se puede abrir una nueva tarea sin finalizar otra.

De esta manera, se pretende dar respuesta al problema habitual de muchas empresas de tener muchas tareas abiertas pero con un ratio de finalización muy bajo. Aquí lo importante es que las tareas que se abran se cierren antes de empezar con la siguiente.

### 2.3.4 Control del flujo

A diferencia de SCRUM, la metodología Kanban no se aplica a un único proyecto, sino que mezcla tareas y proyectos. Se trata de mantener a los trabajadores con un flujo de trabajo constante, las tareas más importantes en cola para ser desarrolladas y un seguimiento pasivo para no tener que interrumpir al trabajador en cada momento.

Asimismo, dicha metodología de trabajo nos permite hacer un seguimiento del trabajo realizado, almacenando la información que nos proporcionan las tarjetas.

Muchos insisten en destacar las ventajas de Kanban respecto a otras metodología ágiles, como puede ser SCRUM. La posibilidad de poder realizar entregas en cualquier momento, cambiar prioridades al vuelo y la visualización perfecta del flujo, son algunos de los puntos que remarcan como elementos diferenciales y de valor. Sin embargo, no podemos decir que exista una metodología mejor que otra sino que dependerá de la naturaleza de la empresa y la forma de organización de sus procesos internos.

## 2.4 Github

GitHub es una plataforma de desarrollo colaborativo para contener proyectos utilizando el sistema de control de versiones Git. Se utiliza principalmente para la creación de código fuente de programas. El software que opera GitHub fue escrito en Ruby on Rails. Desde enero de 2010, GitHub opera bajo el nombre de GitHub, Inc. Anteriormente era conocida como Logical Awesome LLC. El código de los proyectos alojados en GitHub se almacena típicamente de forma pública.

#### Características

- Wiki para cada proyecto. El término wiki (palabra que proviene del hawaiano wiki, “rápido”) alude al nombre que recibe una comunidad virtual, cuyas páginas son editadas directamente desde el navegador, donde los mismos usuarios crean, modifican, corrigen o eliminan contenidos que se comparten. Aquí se comparten términos que se usan en la gerga del negocio de zapatos, dado que son términos propio del negocio hace necesario crear la “wiki” para entender y manejar los términos adecuadamente.
- Página web para cada proyecto.
- Gráfico para ver cómo los desarrolladores trabajan en sus repositorios y bifurcaciones del proyecto.
- Funcionalidades como si se tratase de una red social, por ejemplo, seguidores.
- Herramienta para trabajo colaborativo entre programadores.
- Gestor de proyectos de estilo Kanban.



Repositorio Github del proyecto final de máster [https://github.com/jcjimenezb123/PFM\\_PronosticoDemanda](https://github.com/jcjimenezb123/PFM_PronosticoDemanda)







### 3. Planificación del trabajo

Para lograr los objetivos específicos se deben ejecutar las siguientes actividades:

1. Generar DataSets con información necesaria para cumplir los objetivos específicos.
  - Venta, existencia y precio por tienda-día-modelo-talla.
  - Segmentación, catálogo y página de cada modelo.
  - Imágenes de artículos.
2. Limpieza de datos
  - Actualizar valores negativos de existencia, venta o precio a cero.
  - Estimar la venta de días donde no hubo existencia con algún modelo de ML.
3. Proponer y probar las variables que influyen o determinan el comportamiento de la venta.
  - Construir la matriz de correlación con las variables que se tienen para identificar numéricamente su participación en la venta.
  - Probar la incorporación de cada variable y determinar la mejora de la precisión.
4. Construir modelo de clasificación de artículos buenos regulares y riesgo.
  - Aplicar un modelo de clasificación para determinar cuántos grupos se pueden identificar en una misma familia con respecto a sus ventas.
  - Una vez encontrados los grupos, asociar las imágenes de los artículos a su grupo correspondiente.
  - Entrenar un modelo de Machine Learning que clasifique un artículo como Bueno, Regular o Riesgo de acuerdo a sus características de imagen, precio, etc.
  - Estimar las ventas que se tendrá de un artículo de acuerdo a su clasificación y características.
5. Construir la estructura de la red neuronal para hacer el pronóstico de la demanda por artículo.
  - Crear y entrenar un modelo de Series de Tiempo en algún servicio como Amazon, Microsoft Azure o Google para generar el pronóstico de la demanda por cada artículo en un horizonte de 6 a 8 meses.
6. Reportar resultados.
  - Reportar el grado de precisión de cada modelo con el dataset de test.

Las actividades se registran en la sección del proyecto de Github y se van pasando por las distintas fases de tal manera que todos los integrantes del proyecto son enterados del estado de cada tarea.





## 4. Selección de los modelos

La selección de modelos de Machine Learning está orientado a conseguir la solución de cada objetivo del presente proyecto haciendo uso de las capacidades de IA.

### 4.1 Datos

Antes que los modelos puedan ser planteados se requiere de los datos. Los datos forman parte de las entradas que requieren los modelos tanto supervisados como no supervisados así que la selección de las variables o “features” son muy importantes para obtener información útil a partir de ellos.

Las principales variables que se identifican para los diferentes objetivos son:

- Ventas
- Existencia
- Imágenes
- Precios
- Segmentación
- Catálogos
- Tiendas
- Tallas

En periodos donde los catálogos son vigentes.

#### 4.1.1 Generar Datasets

Se seleccionan los catálogos de:

- Botas
- Sandalias
- Confort

- Caballeros

Los cuales son los más representativos ya que generan las ventas más altas de la empresa y tienen temporalidades muy marcadas.

Se generan los datasets de Existencia-Ventas con la estructura: tienda, id, talla ini, talla fin, talla, fecha, existencia, venta.

#### 4.1.2 Carga inicial de datos

Para iniciar nuestro análisis se han realizado extracciones de los sistemas de PriceShoes y con ellas se generaron archivos en formato CSV.

Estos archivos se cargaron en una base de datos SQL Server en una máquina virtual en Microsoft Azure. Es posible conectarse a esta base de datos desde otras plataformas para extraer la información que requieran los modelos a desarrollar.

Inicialmente se ha creado un modelo muy sencillo -imagen siguiente que incluye:

- Información de ventas, incluyendo cantidades vendidas, precios y existencias para cada modelo y fecha analizada. El rango de fechas va del 01-03-2019 al 29-02-2020. Hemos excluido las fechas más recientes por el efecto que la pandemia ha tenido en el comportamiento de los datos, que es claramente atípico.
- Modelos de los distintos productos incluyendo rangos de tallas y tallas específicas, así como su ubicación en los distintos catálogos.
- Lista de tiendas.



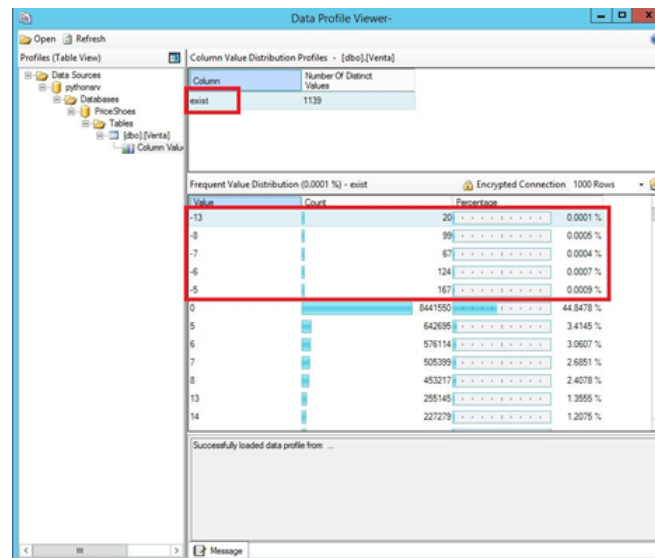
Esta estructura y las capacidades de una base de datos relacional permitirá realizar fácilmente las agrupaciones de datos que se requieran, así como analizar tendencias dentro de los mismos que permitan definir o ajustar los modelos más adecuados para los pasos siguientes.

### 4.1.3 Limpieza de Datos

Utilizando las herramientas de Inteligencia Empresarial que provee SQL Server se ha realizado un análisis de los datos cargados para determinar si existen alguna anomalía dentro de ellos. Se utilizó la herramienta “Data Profiling Task” dentro de las facilidades de SQL Server Integration Services.

En general la calidad de los datos es alta, lo cual representa una gran ventaja. Si se detectaron algunos datos de existencias negativas, que pueden ser fácilmente corregidos y algunas agrupaciones adicionales por proveedor que deben hacerse.

En la imagen siguiente se muestra parte del análisis realizado con el “Data Profiling Task” que permitió la detección de existencias negativas, aunque son muy pocos casos respecto al total.



En los pasos siguientes se realizarán estimaciones de ventas para los casos donde la existencia ha sido cero, es decir aun teniendo demanda por el producto esta no ha podido ser atendida. Tener existencia no garantiza una venta en todos los casos obviamente, pero es un hecho que algunas ventas si se pierden por esta causa y es justo lo que buscamos estimar.

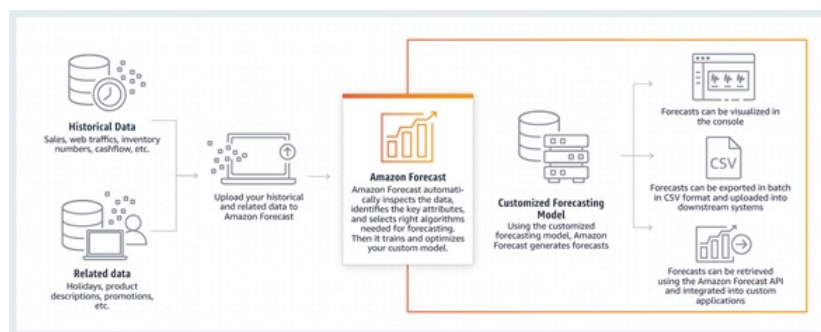
## 4.2 Pronóstico de la demanda

A partir de los datos de la tienda número 1 se procedió a crear un dataset que pueda utilizarse en sistema de machine learning comerciales que es el caso de Amazon Forecast el cual brinda las siguientes cualidades:

“Amazon Forecast incluye algoritmos basados en más de veinte años de experiencia en pronósticos y experiencia desarrollada utilizada por Amazon.com. Amazon Forecast seleccionará automáticamente el mejor algoritmo en función de sus conjuntos de datos específicos. Amazon Forecast puede usar prácticamente cualquier información histórica de series de tiempo (por ejemplo,



precio, promociones, métricas de rendimiento económico) para crear pronósticos precisos para su negocio. Por ejemplo, en un escenario minorista, Amazon Forecast utiliza el aprendizaje automático para procesar sus datos de series temporales (como el precio, las promociones y el tráfico de la tienda) y los combina con los datos asociados (como las características del producto, la ubicación del piso y las ubicaciones de las tiendas) para determinar las complejas relaciones entre ellos. Al combinar datos de series temporales con variables adicionales, Amazon Forecast puede ser un 50 % más preciso que las herramientas de pronóstico que no son de aprendizaje automático.”



Para lograr el entrenamiento con este sistema se realizaron los siguientes cambios en la configuración del dataset.

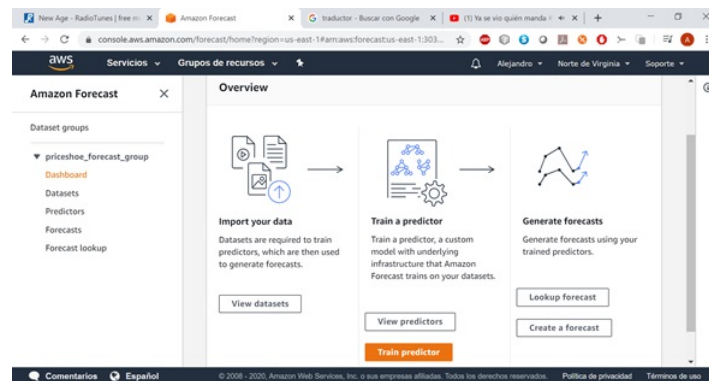
El dataset de ventas original tiene el siguiente formato 1,1560,25,32,25,01/09/2019 00:00:00,5,0 donde el primer número es la tienda, el segundo es el identificador del último sustituto, el tercer formato

Columna	Dato	Descripción
1	1	Número de tienda
2	1560	Artículo
3	25	talla inicial
4	32	talla final
5	25	talla
6	01/09/2019 00:00:00	Fecha
7	5	Invnetario
8	0	Venta

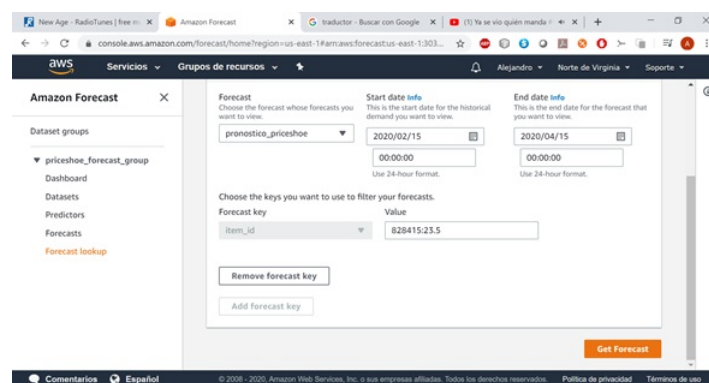
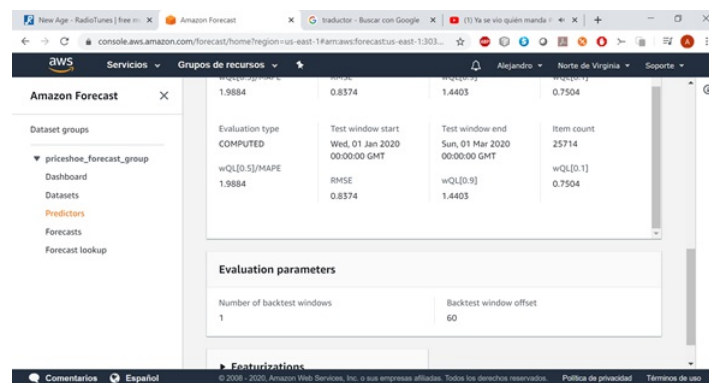
Este formato se tiene que convertir en el siguiente dataset para poder ser entrenado en el servicio Machine Learning Amazon forecast

Columna	Dato	Descripción
1	1560:25	Artículo - talla
2	2019-09-01	Fecha
3	0	Venta

Se tiene que realizar los siguientes pasos para poder generar un pronóstico el cual después puede ser exportado en formato .csv. Cada uno de los pasos tarda aproximadamente 10 minutos para entrenar un predictor y generar el pronóstico.

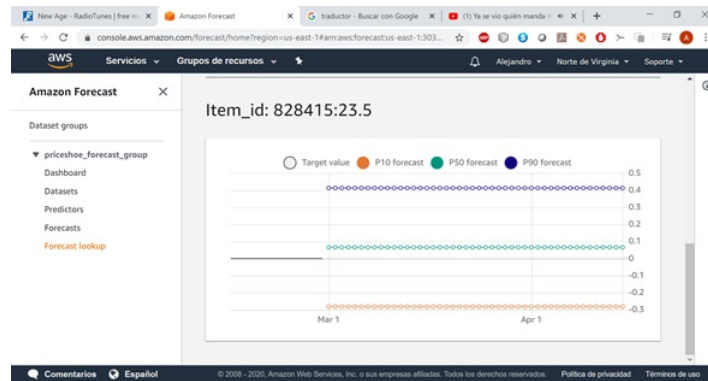


Al realizar las predicciones se estableció una venta diaria de 60 por lo cual se pretende realizar el pronóstico a partir de 28 de febrero de 2020, el data set tiene los datos desde el 1 de noviembre hasta el 28 de febrero de 2020

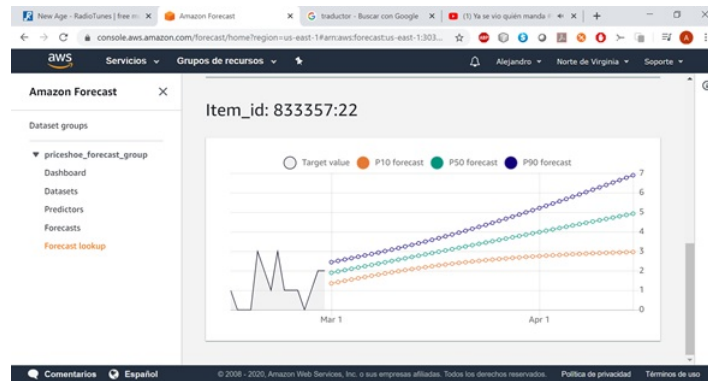


Esta es la vista del pronóstico desde el 15 de febrero al 15 de abril el cual a partir del 28 de febrero presenta un pronóstico con 3 tipos de percentil P10, P50, P90. Cada punto representa la fracción de pares que se venden por día para un periodo de 60 días.

Esta es la vista utilizando el modelo 828415 con número 23.5 es el siguiente:



Esta es la vista utilizando el modelo 833357 con número 22 es el siguiente:





## 5. Construcción y entrenamiento de los modelos







## **6. Métricas y evaluación de los modelos**







## 7. Resultados







## **8. Discusión**







## **9. Conclusiones**