



# Tecnológico de Monterrey

## **Laboratorio de diseño y optimización de operaciones**

Modelos de Aprendizaje de Máquina para predicción de  
demanda de productos de telefonía celular Huawei

Ana Luisa Masetto Herrera

**Presentan:**

Brenda Aurora Rubio Villaseñor A01364693  
María Fernanda Gutiérrez Mujica A01367761  
Rafael Loyola Franco A01365017  
Carlos Muñoz Villegas A01363608

02 de junio, 2021

## **Etapas 1 - Comprensión del Negocio:**

Desarrollar un proyecto de ciencia de datos enfocado a resolver un problema de construcción de portafolios de productos (predicción de demanda), de una marca en específico, para los diferentes puntos de venta de la empresa Huawei, empresa de la industria de telecomunicaciones, enfocada al sector de la telefonía celular.

### **1. Descripción de la situación actual**

Huawei es una empresa de tecnología china multinacional, con sede en Shenzhen inicialmente centrada en la fabricación de computadoras y teléfonos ha expandido su gama de productos, está presente en más de 170 países, es el segundo mayor fabricante de teléfonos inteligentes.

Una de las noticias más importantes para Huawei fue cuando, debido a la guerra comercial entre EUA y CH el ex presidente Donald Trump prohibió que la marca hiciera negocios con empresas norteamericanas e hizo un bloqueo comercial, otro factor que influye en esta decisión de parte del gobierno americano es el supuesto espionaje a organismos clave de la seguridad de EUA de parte de China a través de la arquitectura de telecomunicaciones de Huawei.

Huawei se enfoca en soluciones integradas en cuatro áreas clave: redes de telecomunicaciones, tecnología de información, dispositivos inteligentes y servicios de nube están comprometidos a llevar lo digital a cada persona, hogar y organización para un mundo inteligente y totalmente conectado.

Huawei ha logrado ganar la confianza de los mexicanos, lo cual se ve en los resultados que ha tenido la marca en México, ya que se ha logrado posicionar como la segunda marca más vendida de smartphones después de el fabricante asiático Samsung, con un 16% total de dispositivos colocados en el mercado.

Creemos que este proyecto es importante dado que hoy en día en un mundo tan globalizado y con tanta competencia no se pueden tomar decisiones que no estén basadas en datos, es necesario tener argumentos sólidos y fundamentados para que de esta manera se logre incurrir en los menores costos posibles, estos si es que se pretende seguir siendo competitivo.

Para este proyecto en particular creemos que es necesario entender los patrones de consumo de los distintos consumidores a través de los puntos de venta y en las diferentes regiones, de esta manera tener un mejor entendimiento de lo que el cliente necesita y cuáles son sus prioridades, para así ofrecer los productos indicados que mejor cubran esas necesidades y generar el máximo aprovechamiento de los mismos con el objetivo de lograr la satisfacción del cliente y disminuir nuestra estructura de costos.

### **2. Entender y describir la problemática (en términos del negocio)**

Al ser una empresa que cada día crece más y el consumo de sus productos es cada vez más alto, Huawei se enfrenta a retos grandes en todos los aspectos del negocio. Principalmente usar toda su fuerza creativa y tecnológica para desarrollar nuevos productos y actualizar aquellos que ya están en venta. El crear un sistema operativo es un gran reto en el que han estado trabajando para sus propios equipos desde 2017 y que se espera ser lanzado próximamente. Además de los problemas diarios que presenta la cadena de suministro, este punto es muy importante ya que es una de las razones más grandes por las cuales están entre las mejores empresas dentro de la tecnología. Para seguir teniendo el éxito que los ha colocado tan alto en la popularidad de los consumidores, es prioridad trabajar de la mano con todos sus puntos de venta en las distintas regiones donde están presentes, contar con

ingenieros capacitados para interactuar con las distintas conexiones de la cadena de suministro.

**3. Entender y describir la problemática (en términos de ciencia de datos, tipo de tarea, tipo de datos, etc.)**

Gracias a la minería y gestión de datos es que se han podido apoyar para contar con un sistema que les permita abastecer y pronosticar las ventas de las distintas regiones en los diferentes periodos del año. Un problema grande que tienen a comparación de otras empresas como Apple, es que Huawei cuenta con teléfonos de diferentes precios, además de vender televisores, computadoras, tabletas, sistemas de audio y más. Eso convierte el problema a uno más complejo en donde los datos y sus lecturas juegan un papel crucial.

Nuestro objetivo es realizar una predicción de la demanda sobre los distintos puntos de venta, es decir, no solo responder a la predicción sobre la demanda sino qué productos y puntos de venta adquieren relevancia a partir de esta predicción.

Esto lo convierte en un problema de regresión con el cual se desea ajustar un modelo el cual pueda predecir con cierto grado de exactitud y certeza la demanda de las distintas regiones y puntos de venta para los productos.

La pregunta a responder es ¿Cuántas unidades de cada producto de mi marca, se van a vender en los diferentes puntos de venta, al siguiente mes de registro?

**4. Plasmar los objetivos**

Según información de la página oficial de la empresa, el objetivo principal es ayudar a nuestros clientes y socios a ser más competitivos, aumentar sus ingresos o reducir sus costes, y permitirles alcanzar un mayor éxito empresarial. En el pasado, Huawei hizo muchas innovaciones de ingeniería y tecnología en redes inalámbricas, ópticas y dispositivos inteligentes. Estas innovaciones generaron una gran cantidad de valor comercial para nuestros clientes y un enorme valor social.

Objetivos del proyecto:

- a. Comprender e investigar la situación en la que se encuentra la empresa, la problemática y los objetivos de esta para así poder estructurar de forma adecuada el plan preliminar del proyecto.
- b. Comprender y detectar los problemas de calidad con respecto a los datos proporcionados.
- c. Construir modelos que permitan obtener la solución del problema.
- d. Evaluar los modelos y concluir con respecto a su desempeño.

**5. Estructurar el proyecto y hacer un plan preliminar**

- Business understanding
  - Entender los datos
  - Establecer objetivos y requerimientos desde el punto de Huawei para pronosticar las ventas y tener un mejor modelo de negocio
- Data understanding
  - Hacer un reporte con la descripción de los datos
  - Explorar los datos
  - Hacer un reporte de la calidad de los datos
- Data preparation
  - Seleccionar los datos
  - Limpiar los datos
- Modeling
  - Aplicar varios modelos con parámetros predefinidos

- Justificar la selección de los modelos elegidos
- Tener en cuenta que hay técnicas que tienen requerimientos especiales en el formato de los datos
- Evaluation
  - Evaluar los resultados
  - Revisión del proceso
  - Determinar los siguientes pasos
- Deployment
  - Organizar la información y los resultados
  - Monitoreo y mantenimiento del plan
  - Reporte final
  - Revisión del proyecto
  - Presentación de resultados para que el cliente pueda usar la información obtenida

## **Etapas 2 - Comprensión de los datos:**

### **1. Describir los datos crudos**

Se tienen 14 variables y 369,617 datos en total en la base de datos de Huawei, y la terminología es la siguiente:

<b>Variable</b>	<b>Tipo de Dato</b>	<b>Definición</b>	<b>Problema de calidad</b>
punto_de_venta	Caracter	Punto de contacto del usuario con la marca Huawei para comprar.	2 puntos de venta mal escritos. Debe corregirse.
fecha	Caracter	Fecha de cuando se realizó la venta (día, mes y año).	No requiere correcciones.
mes	Caracter	Mes en el que se realizó la venta.	Hay 6 meses escritos con letras mayúsculas en lugar de números. Debe corregirse.
anio	Entero	Año en el que se realizó la venta.	Un año (2018) escrito en otra notación. Debe corregirse.
num_ventas	Entero	Cantidad de ventas realizadas en ese día.	No requiere correcciones.
sku	Caracter	Número de referencia único para el celular vendido.	No requiere correcciones.
marca	Caracter	Marca de teléfono	Huawei tiene 4

		móvil vendido.	alteraciones del nombre. Se debe corregir.
gamma	Caracter	Térmico establecido de acuerdo al precio del celular.	No requiere correcciones.
costo_promedio	Numérico	Costo aproximado en el que incurrió Huawei en la producción del celular vendido.	Un dato atípico. Requiere inspección.
zona	Caracter	Zona en la que fue vendido el celular como centro sur, centro occidente, golfo de México, etc.	Una zona repetida con diferente escritura. Se debe corregir.
estado	Caracter	Estado de la República Mexicana donde fueron vendidos los equipos.	“La Paz” aparece además de Baja California. Se debe corregir.
ciudad	Caracter	Ciudad de la República Mexicana donde fueron vendidos los equipos.	No requiere correcciones.
latitud	Numérico	Coordenadas geográficas de venta.	Un valor fuera de rango. Se debe corregir.
longitud	Numérico	Coordenadas geográficas de venta.	Un valor fuera de rango. Se debe corregir.

## 2. Detectar problemas de calidad (solamente enunciar los problemas, en la siguiente etapa se menciona cómo se hizo la limpieza)

Al hacer una revisión de la base de datos crudos, se puede observar que hay irregularidades en la forma en la que los datos fueron ingresados de ciertas variables tales como: punto de venta, mes, año, marca, zona, estado, latitud y longitud. Estas irregularidades o variantes en los nombres puede causar que se esté hablando de la misma cosa pero la base de datos las tome por dos distintas. Es por eso que se debe realizar una limpieza de datos para reducir el error a la hora de utilizar dichos datos y poder continuar con los objetivos establecidos en el proyecto.

### **Etapas 3 - Data Preparation**

- **Limpieza de datos**

Para poder analizar los datos fue necesario hacer una limpieza de las variables ingresadas incorrectamente. Al hacer una revisión extensa de los datos pudimos observar que en algunas variables había algunos datos que fueron ingresados de la forma incorrecta lo cual haría imposible trabajar con esta base de datos en el futuro.

En la variable punto de venta se encontraron 5 diferentes puntos que fueron ingresados de forma incorrecta los cuales son y se corrigieron de la siguiente manera:

- ace alhondga- ace alhondiga
- ACE MEGA iguala - ace mega iguala
- acr centroapizacotlx - acr centroapizacotl
- acr CENTROapizacotl - acr centroapizacotl
- mgn aleira gdl - mgn alegra gdl

En la variable de mes hubo un error al ingresar los datos. En algunas ocasiones se ingresó la abreviatura del mes como por ejemplo: SEPT y en otras ocasiones se ingresó de forma correcta como es el número que corresponde a cada mes. Fueron corregidas de la siguiente forma:

- JUN-7
- JUL-7
- SEPT-9
- DIC-12
- MAR -3

La variable año debe de ser ingresada con el formato de 4 dígitos. Esta variable fue modificada de la siguiente manera: "18" "2018"

En la variable marca se encontraron 5 marcas que no fueron ingresadas de la forma correcta las cuales fueron:

- Huawei-huawei
- HUAWEI-huawei
- hHuawei- huawei
- hhuawei-huawei

En la variable de Zona hay una variable que no fue ingresada de forma correcta la cual es: CENTRO SUR- centro sur

La variable estado tampoco fue ingresada de forma correcta. Existen tres registros más de estados las cuales son:

- "merida" "yucatan"
- "la paz" "baja califronia sur"
- "metepec" "estado de mexico"

Tanto en las variables de longitud y latitud se encontraron 2 valores los cuales fueron encontrados fuera de rango y fueron corregidos de la siguiente manera:

Longitud  
-9949052-99.49052  
Latitud  
1941515-19.41515

- **Comportamiento de datos en el tiempo (Gráficas)**

Hicimos un análisis de los datos y realizamos varias gráficas que nos pudieran mostrar un poco mejor el panorama y comportamiento de los datos de Huawei, los resultados se muestran a continuación:

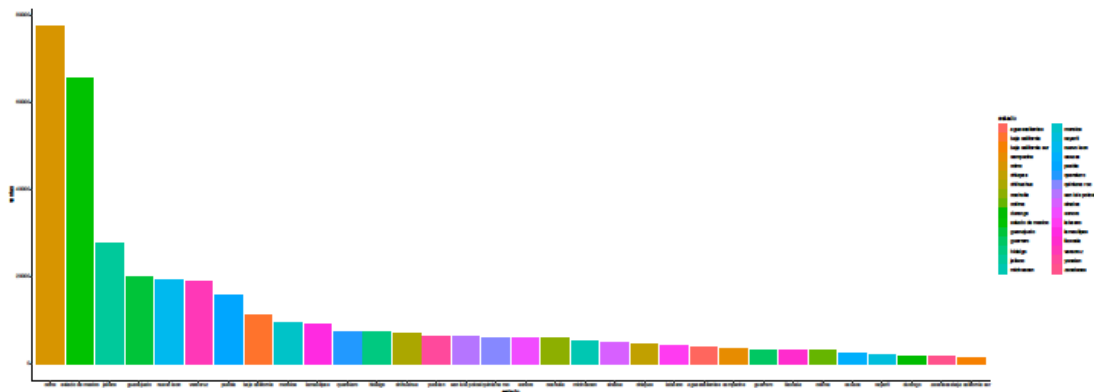


Gráfico 1. Ventas por estado

Gracias a esta gráfica podemos identificar visualmente que el mayor número de ventas se concentra en la Ciudad de México, seguido por el Estado de México y en tercer lugar se encuentra Jalisco. Guanajuato, Nuevo León y Veracruz le siguen y se puede ver que tienen un comportamiento de ventas bastante similar en cuanto a cantidad.

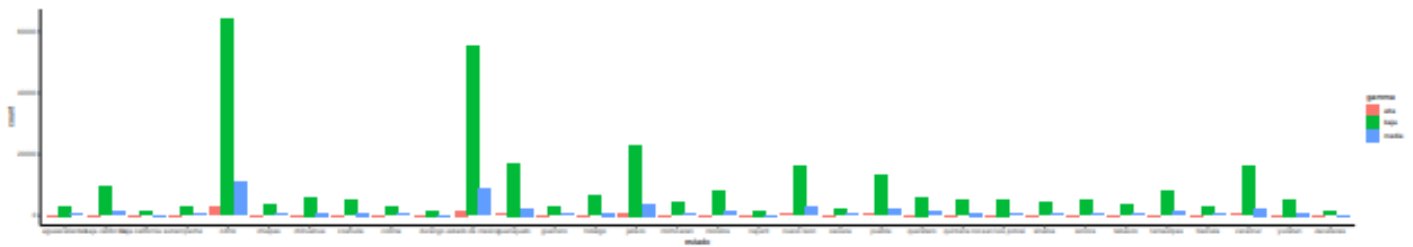


Gráfico 2. Gamma por estado

En esta gráfica nos concentramos en los estados de mayor relevancia en cuanto a número de ventas que pudimos identificar en la primera parte. Los 6 estados mencionados anteriormente, a pesar de ser los que tienen más ventas, se puede observar cómo la mayor parte de éstas se concentran en modelos de gamma baja, seguido por la media y al último la alta.

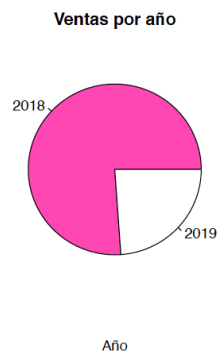


Gráfico 3. Ventas por año

Si bien tenemos datos de las ventas de 6 meses del 2018 vs 3 meses de 2019, gracias esta gráfica de pastel podemos identificar que las ventas de 2018 a 2019 se redujeron drásticamente pues de los dos años, el 2018 abarca un poco más de 3/4 del total de ventas cuando se esperaría que ocupara aproximadamente un 66%. Algún factor externo que pueda estar jugando en esta parte es que 2018 abarca Diciembre que es un mes que suele ser fuerte por las festividades en que se llegan a comprar varios celulares.

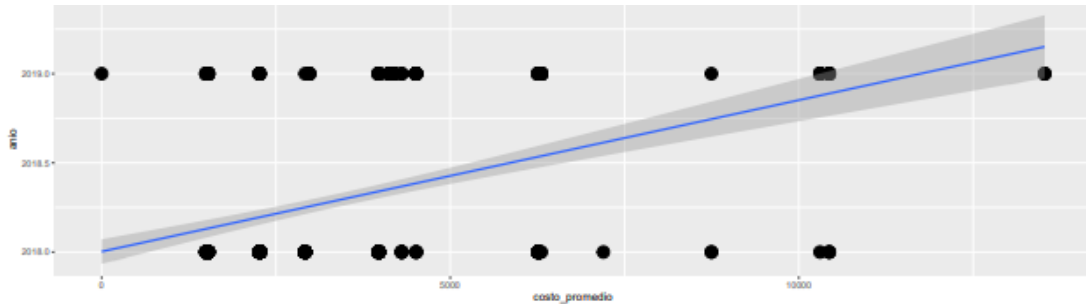


Gráfico 4. Gráfico de dispersión

En esta gráfica podemos ver la tendencia que ha seguido el costo promedio a lo largo de los años, esto de los últimos 500 datos, como éstos poseen una tendencia a la alza y una correlación positiva dado que conforme va a aumentado la variable de año y pasa el tiempo, de igual manera el costo promedio de los productos aumenta.

- **Seleccionar y construir ingeniería de características**

Haciendo uso de diversas funciones como lo son join, merge y left\_join las cuales ayudan a juntar y/o relacionar datos fuimos capaces de cambiar nuestras variables categóricas, otorgándoles un índice y de esta manera relacionarlas con otros datos con el objetivo de generar nuevas variables que puedan ser utilizadas y que aporten mucho más valor y una perspectiva diferente, para así obtener conclusiones más enriquecedoras de nuestros datos para el proyecto.

Nuestro dataset tiene 581,094 observaciones de 24 variables.

Con la incorporación de las siguientes variables:

- Ventas totales
- Ventas totales por tienda, por mes y el acumulado de dos y tres meses pasados
- Ventas totales por tienda y por SKU
- Ventas promedio por tienda, por mes y el acumulado de dos y tres meses pasados
- Ventas promedio por SKU, por mes y el acumulado de dos y tres meses pasados
- Ventas promedio por tienda y por SKU de dos y tres meses pasados
- Ventas del siguiente mes

Estas variables son importantes a considerar debido a que dan una imagen mucho más completa de como es el comportamiento de las ventas de los productos, lo cual es el objetivo del proyecto de esta manera podemos analizar estos patrones y obtener pronósticos mucho más acertados de la demanda de los celulares y insights valiosos de utilidad para la compañía.



## Etapa 4 - Modelado

### Promedios móviles

El promedio móvil es un indicador de tendencias que se usan para realizar análisis de datos anteriores con la finalidad de formar una serie de medidas que provengan de diversos subconjuntos de datos, por lo tanto, tienen la capacidad de examinar las medidas de datos que disminuyen o aumentan en un periodo de tiempo. Estas medidas se pueden clasificar como conductores, ya que utilizan diversos datos anteriores para trazar nuevas líneas de datos. Se manifiestan por medio de señales luego de que la nueva tendencia o los movimientos importantes de datos hayan comenzado.

### Consideraciones

Las series de tiempo horizontales son aquellas gráficas que representan a un proceso que se encuentra estable en el tiempo. La variación que existe se debe únicamente a la variación natural. Cuando se cuenta con este tipo de gráfica se puede emplear el método para realizar los pronósticos de promedio móvil. Al estar disponible cada nueva observación, se puede calcular una nueva media eliminando el valor más antiguo e incluyendo el más reciente. Entonces, se usa este promedio móvil para pronosticar el siguiente periodo.

### Limitantes del Promedio Móvil

En los promedios móviles solamente es posible llegar a pronosticar un período más. Suelen ser simplificaciones reales y no garantizan las variables influyentes en el futuro de los pronósticos que se encuentren incluidos en el modelo de dicho pronóstico.

Al utilizar el método promedios móviles con los datos de Huawei, quisimos comparar 3 opciones:

1. Utilizando el periodo anterior
2. Obteniendo el promedio de los 2 meses anteriores
3. Obteniendo el promedio de los 3 meses anteriores

```
In [10]: datos_ma['m1_pedir_mes_anterior'] = datos_ma['ventas_totales']
```

```
In [11]: datos_ma.head()
```

```
Out[11]:
```

	pdv_id	mes_id	sku_id	ventas_totales	y_ventas_siguiente_mes	m1_pedir_mes_anterior
0	1	0	1	1	0	1
1	1	1	1	0	0	0
2	1	2	1	0	0	0
3	1	3	1	0	0	0
4	1	4	1	0	0	0

#### 4.2 Pedir el promedio de 2 meses anteriores

```
In [12]: datos_ma['m2_pedir_2_meses_anteriores'] = datos_ma.groupby(['pdv_id', 'sku_id']).rolling(2)['ventas_totales'].mean().reset_index()
```

```
In [13]: datos_ma.head()
```

```
Out[13]:
```

	pdv_id	mes_id	sku_id	ventas_totales	y_ventas_siguiente_mes	m1_pedir_mes_anterior	m2_pedir_2_meses_anteriores
0	1	0	1	1	0	1	NaN
1	1	1	1	0	0	0	0.5
2	1	2	1	0	0	0	0.0
3	1	3	1	0	0	0	0.0
4	1	4	1	0	0	0	0.0

#### 4.3 Pedir el promedio de 3 meses anteriores

```
In [14]: datos_ma['m3_pedir_3_meses_anteriores'] = datos_ma.groupby(['pdv_id', 'sku_id']).rolling(3)['ventas_totales'].mean().reset_index()
```

```
In [15]: datos_ma.head()
```

```
Out[15]:
```

	pdv_id	mes_id	sku_id	ventas_totales	y_ventas_siguiente_mes	m1_pedir_mes_anterior	m2_pedir_2_meses_anteriores	m3_pedir_3_meses_anteriores
0	1	0	1	1	0	1	NaN	NaN
1	1	1	1	0	0	0	0.5	NaN
2	1	2	1	0	0	0	0.0	0.333333
3	1	3	1	0	0	0	0.0	0.000000
4	1	4	1	0	0	0	0.0	0.000000

Para poder identificar qué método resultaba mejor, sacamos el MAE (Mean Absolute Error) que es el promedio absoluto del error:

	Mes	mae_pedir_anterior	mae_promedio_2_meses_anteriores	mae_promedio_3_meses_anteriores
0	Julio	0.568813	NaN	NaN
1	Agosto	0.488399	0.528312	NaN
2	Septiembre	0.428352	0.463897	0.508312
3	Octubre	0.477883	0.508650	0.557967
4	Noviembre	0.697798	0.704496	0.729977
5	Diciembre	0.740994	0.656135	0.665644
6	Enero	0.645340	0.714982	0.672113
7	Febrero	0.514838	0.600401	0.701195
8	Marzo	0.518307	0.523828	0.570434

Tabla 1. Cálculo de MAE

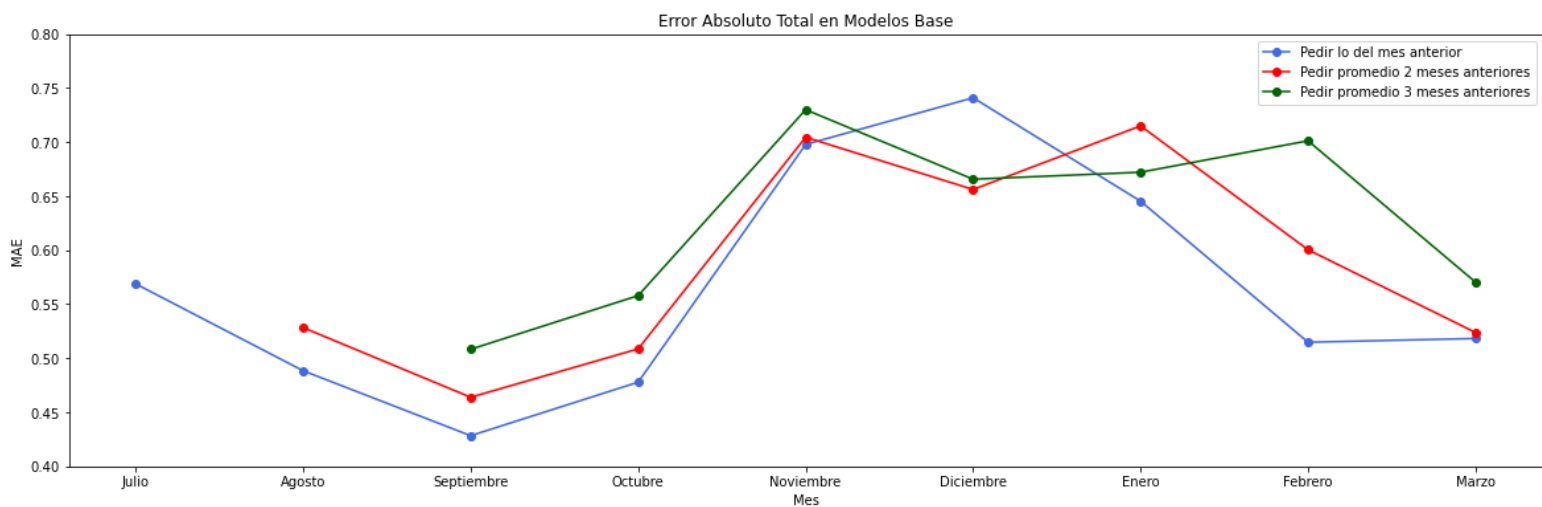


Gráfico 5. Error absoluto en modelos base

## Resultados de desempeño que obtuvimos

Creemos que el mejor modelo a utilizar es el mae en el que se pronostica de acuerdo a las ventas del periodo anterior ya que tiene el menor error promedio en todos los meses salvo Diciembre, pero, si se evalúan los 9 meses en conjunto, sí nos conviene la utilización de este método.

En Diciembre observamos un pico con el método elegido ya que podemos deducir que las ventas en este mes incrementaron a causa de las festividades como Navidad y Año Nuevo. En Enero disminuye drásticamente el error ya que podemos intuir que Huawei continuó con ventas altas por ventas post año nuevo y día de reyes. Pasado este mes, el error se empieza a regularizar.

- **Modelo de aprendizaje de máquina**

El modelo de aprendizaje de máquina que se eligió y desarrolló para este caso fue el de: árboles de decisión, que consiste en un tipo de algoritmo de aprendizaje supervisado y se usa principalmente en problemas de regresión.

Dentro de las principales ventajas de los árboles de decisión encontramos que llegan a ser útiles con o sin datos fehacientes, sin embargo, dentro de las principales ventajas que tiene se encuentran:

- Pueden ser útiles con o sin datos fehacientes, es decir, que las variables de entrada y salida pueden ser en especie continuas o categóricas.
- Divide el espacio de predictores en regiones no sobrepuestas y diferentes.

Entre las limitantes que caracterizan a este modelo, se encuentran las siguientes:

- Pérdida de información al catalogar variables continuas.
- El desarrollar divisiones estratégicas puede alterar la precisión del árbol.
- Se puede modificar la estructura del árbol de manera muy significativa, con un pequeño cambio de datos, suele presentarse un poco de inestabilidad.
- Las máquinas de vectores de soporte, que es otro tipo de modelo de aprendizaje de máquina, generalmente tienen tasas de error 30% más bajas que los árboles de decisión.

## Resultados de desempeño

Conjunto	Mes	dt_1_profundidad	dt_5_profundidad
entrenamiento	julio	0,5778	0,4695
entrenamiento	agosto	0,5239	0,4397
entrenamiento	septiembre	0,4909	0,4229
entrenamiento	octubre	0,4828	0,4297
entrenamiento	noviembre	0,5336	0,4667
entrenamiento	diciembre	0,5547	0,5086
entrenamiento	enero	0,5522	0,0527
entrenamiento	febrero	0,5424	0,54
prueba	julio	0,5985	0,5382
prueba	agosto	0,4251	0,4155

prueba	septiembre	0,4583	0,4515
prueba	octubre	0,7369	0,6867
prueba	noviembre	0,6604	0,9001
prueba	diciembre	0,5373	0,6104
prueba	enero	0,0474	0,5092
prueba	febrero	0,3651	0,4835

Tabla 2. MAE de modelos de aprendizaje de máquina

Para medir el desempeño de nuestros diferentes modelos de aprendizaje de máquina fue necesario calcular el MAE y con ello determinar cuál era el modelo que realizaba el pronóstico con la menor cantidad de error posible y nos entregará un mejor balance, es por ello que determinamos que el mejor modelo es el de Árboles de decisión de profundidad 1.

### Etap 5 - Evaluación (Resultados - Conclusiones)

Decidimos realizar la comparación de los errores absolutos medios de promedios móviles y árboles de decisión con profundidad 1 y 5. Respecto al primer método empleado, que es el de promedios móviles, decidimos realizar la comparación utilizando el dato del mes anterior ya que una vez efectuados los promedios utilizando 2 y 3 meses anteriores, sus indicadores de desempeño resultaron más altos a comparación de éste.

En la siguiente gráfica podemos observar los errores absolutos medios de los modelos antes mencionados.

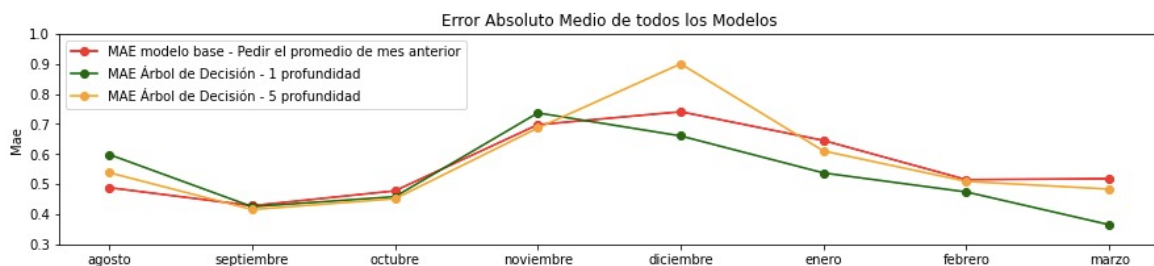


Gráfico 6. Error absoluto medio de los 3 modelos empleados

Decidimos que el mejor modelo a emplear es el de árboles de decisión con profundidad 1. Si bien de agosto a noviembre no es el modelo con el menor error absoluto medio, de diciembre a marzo es el que nos proporciona significativamente los mejores pronósticos.

Modelo de aprendizaje de máquina					
pdv_id	mes_id	sku_id	ventas_totales	y_ventas_siguiente_mes	venta_por_mes_pred
82	4	12	2	2	0
		13	5	5	5
		14	1	0	0
		15	2	0	0
		16	3	0	0

Tabla 2. Pronóstico de ventas para Diciembre con modelos de ML.

Una vez decidido el modelo de pronóstico, podemos entonces ya identificar por mes, sku de producto y punto de venta, los pronósticos de unidades a vender para el mes que sigue. En la tabla 2 podemos observar que para el mes de Diciembre, se venden 7 unidades en total de los artículos con sku 12-16 en el punto de venta con ID #82 y nuestro modelo hizo un pronóstico de venta para Diciembre de 5 artículos en total para los mismos sku.

Promedios móviles				
pdv_id	mes_id	sku_id	y_ventas_siguiente_mes	m1 pedir mes anterior
82	4	12	2	2
82	4	13	5	5
		14	0	1
		15	0	2
		16	0	3

Tabla 3. Pronóstico de ventas para Diciembre con modelo de Promedios Móviles.

En este modelo de promedio móviles podemos detectar que al comparar los mismos datos en la tabla 3 podemos observar que para el mes de Diciembre , de igual manera se venden 9 unidades en total de los artículos con sku 1,2,3,4,5 en el punto de venta con ID #1 y nuestro modelo hizo un pronóstico de 4 artículos en total para los mismo sku.

Al final del día, cualquier método a emplear tendrá su margen de error, sin embargo, habrá que elegir aquel que lo minimice ya que de esta forma podremos tener mejores resultados (pronósticos en este caso). Huawei ahora puede prever qué tipo de unidades tener en stock en sus respectivas sucursales de acuerdo al punto de venta registrado para ofrecer a su cliente el mejor de los servicios y de esta manera tener usuarios satisfechos y generar aún mayores utilidades.

## ANEXO

- **Gantt chart**

TAREAS	RESPONSABLE	INICIO	FINALIZACIÓN	DÍAS	ESTADO
Etapa 1: Business Understanding					
Entender los datos	Todos	1-mar	3-mar	3	Completado
Establecer objetivos y requerimientos del proyecto	Carlos Muñoz	2-mar	5-mar	4	Completado
Etapa 2: Data Understanding					
Hacer un reporte con la descripción de los datos	Rafael Loyola	3-mar	5-mar	3	Completado

Explorar los datos	Fer Gutiérrez	3-mar	6-mar	4	<b>Completado</b>
Hacer un reporte de la calidad de los datos	Brenda Rubio	5-mar	7-mar	3	<b>Completado</b>
Etapa 3: Data preparation					
Seleccionar los datos	Rafael Loyola	3-mar	10-mar	8	<b>Completado</b>
Limpiar los datos	Todos	3-mar	22-mar	20	<b>Completado</b>
Etapa 4: Modeling					
Aplicar varios modelos con parámetros predefinidos	Fer Gutiérrez	1-abr	15-abr	15	<b>Completado</b>
Justificar la selección de los modelos elegidos	Brenda Rubio	15-abr	22-abr	8	<b>Completado</b>
Tener en cuenta que hay técnicas que tienen requerimientos especiales en el formato de los datos	Carlos Muñoz	22-abr	1-may	10	<b>Completado</b>
Etapa 5: Evaluation					
Evaluar los resultados	Todos	3-may	10-may	8	<b>Completado</b>
Revisión del proceso	Rafael Loyola	10-may	14-may	5	<b>Completado</b>
Determinar los siguientes pasos	Todos	14-may	16-may	3	<b>Completado</b>
Etapa 6: Deployment					
Organizar la información y los resultados	Brenda Rubio	16-may	18-may	3	<b>Completado</b>
Monitoreo y mantenimiento del plan	Fer Gutiérrez	18-may	20-may	3	<b>Completado</b>
Reporte final	Carlos Muñoz	20-may	25-may	6	<b>Completado</b>
Revisión del proyecto	Todos	25-may	26-may	2	<b>Completado</b>
Presentación de resultados para que el cliente pueda usar la información obtenida	Todos	26-may	31-may	2	<b>Completado</b>
				<b>87</b>	<b>días</b>