

PRESENTACIÓN

FINAL:APPLE DATA

Equipo 4: DATALENTED

Daniela Monserrat Garcia Sotelo

A01365499

Jorge Abraham Sánchez Mora

A01364653

Introducción

Apple, una de las empresas más exitosas del mundo , productora de equipos electrónicos, software y servicios, es víctima de esta problemática y busca desarrollar la metodología adecuada para satisfacer las demandas netas de los clientes.

Desde el año 2012, Apple empezó a tener un peso muy importante en el mercado mexicano, convirtiéndose en la empresa con el mayor porcentaje de demanda en el país, superando a marcas como Huawei y Samsung. (nótese en la gráfica siguiente).

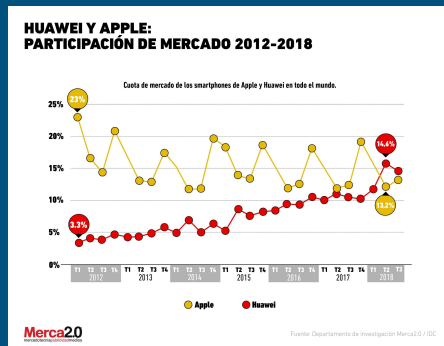
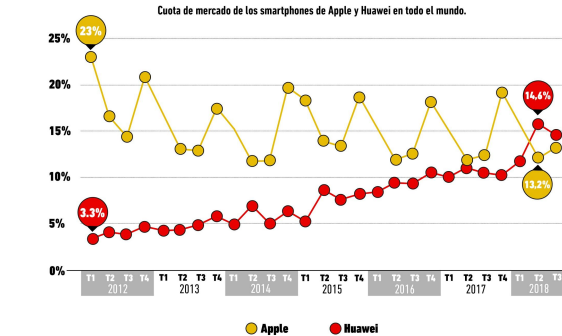


Figura 1. Participación de Mercado (Merca 20, 2018)

HUAWEI Y APPLE: PARTICIPACIÓN DE MERCADO 2012-2018



Merca20

Fuente: Departamento de Investigación Merca20 / IDC

Figura 1. Participación de Mercado
(Merca 20, 2018)

Como se mencionó anteriormente, Apple presenta una de las mayores problemáticas dentro de su industria, la predicción de demanda. Existen diversos factores que propician los cambios drásticos en las necesidades de las personas y su enfoque en la búsqueda de dispositivos celulares que las satisfagan. Actualmente, las personas ya no solo buscan un teléfono cualquiera, se trata de desarrollar un dispositivo que esté al alcance de cualquier usuario, que pueda interactuar con él de manera inteligente.

Etapa 1: Comprensión del negocio



Tenemos la tarea de analizar una gran cantidad de datos referentes a los historiales de demanda de productos Apple en diversos puntos de venta, alrededor de México. Nos enfrentamos a un problema de regresión para pronósticos futuros, además de un conjunto de datos estructurados.

Nuestro conjunto de datos se debe estudiar, limpiar y modificar de manera correcta para lograr un uso adecuado de estos; una vez realizada esta etapa, utilizando la metodología CRISP-DM y herramientas de programación como Rstudio, se propondrá y desarrollará un modelo que nos permita resolver un problema de construcción de portafolios de productos, es decir, predecir las demandas de telefonía celular de la marca Apple, también denominados iPhone, en los próximos meses.



Objetivos

- a) Satisfacer los gustos y necesidades del cliente.
- b) Crear y desarrollar productos innovadores que se adapten a las nuevas tendencias tecnológicas.
- c) Ofrecer un producto que pueda combinar belleza, inteligencia, innovación y accesibilidad al mismo tiempo.



Etapa 2: Comprensión de los datos

1) Describir los datos crudos

Se nos dio una tabla de datos estructurados con 102 530 datos, dentro de los cuales, en cada columna se encuentra la siguiente información:

Puntos de venta: Refieren al lugar dónde se ubican los productos apple.

Fecha: Refiere al día que se entregó el producto.

Año: Es el año en el que el producto se vendió.

Número ventas: Es el número de unidades vendidas que se obtuvo.

Costo promedio: Costo total promedio del producto.

Marca: Refiere a la marca del producto, es decir, el modelo del mismo.

Gamma: Indica el nivel del producto.

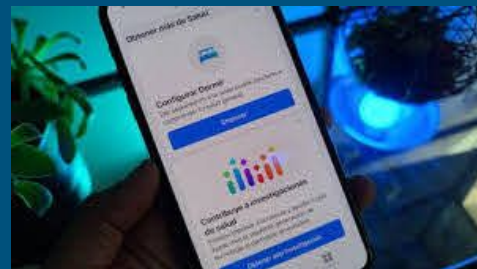
Zona: Zona o lugar donde se vendió el producto.

Estado: Estado, del país, donde se vendió el producto.

Ciudad: Ciudad donde se vendió el producto.

Latitud: Latitud se refiere a aquella donde se encuentra el punto de venta

Longitud: La longitud es el punto geográfico del punto de venta.

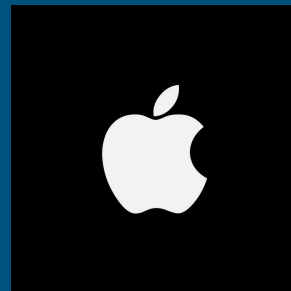


Detectar los problemas de calidad

- **Fecha:** Los registros están limpios.
- **Año:** Debe tomar valores de 4 dígitos
- **Número de ventas:** Los registros están limpios.
- **Puntos de venta:** Existen 5 puntos de venta de manera errónea, ya sea por mayúsculas, acentos, etc.
- **Variable Mes:** Es una variable numérica, hay que cambiarla. Hay 5 valores mal registrados, en lugar de números, hay letras.
- **Variable Sku:** Los registros están limpios.
- **Variable Marca:** Hay 5 marcas escritas de forma errónea.
- **Variables Costo, Ciudad y Gamma:** Registros limpios.
- **Variable Zona:** Hay un registro mal escrito
- **Variable Estado:** Hay tres estados más de los que en realidad existen.
- **Variables Latitud y Longitud:** Hay un valor fuera de rango.



Etapa 3: Preparación de datos



Limpieza de datos

La cantidad de datos que nos fueron proporcionados era muy grande, por lo tanto, analizar estos errores de manera directa y manual nos llevaría demasiado tiempo. Si bien es cierto, nuestro asesor pedagógico menciona que, generalmente, al tratarse de este tipo de proyectos, los analíticos o especialistas en ciencia de datos invierten una gran parte de su tiempo en esta limpieza. Por consiguiente, se utilizó R Studio, una herramienta de programación que nos permitía observar de manera simple este tipo de errores; con simples funciones como “*unique*” o “*select*”, ya éramos capaces de identificar posibles equivocaciones en los datos. Finalmente, empleando diversas funciones y programaciones, logramos limpiar de manera correcta cada uno de los datos, para dar seguimiento al proyecto y tener información cien por ciento verídica y confiable.

Análisis exploratorio de datos

Al finalizar la limpieza de nuestros datos, desarrollamos un análisis exploratorio de estos, el cual permite tener una perspectiva o una visualización de las circunstancias de nuestros datos en la realidad, permitiéndonos entender de manera más profunda la situación.

Aunque no se aprecian todos los nombres de los distintos puntos de venta en la imagen, se puede observar en un primer plano, qué estados cuentan con más tiendas con gran facilidad, por ejemplo, la Ciudad de México y el Estado de México, puesto que son los renglones con más cajas azules.

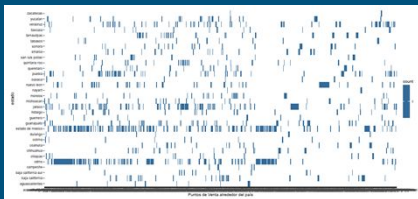
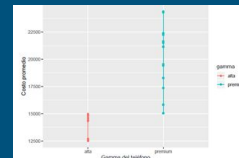


Figura 2. Gráfico Bin2d (Puntos de venta vs. Estados)

Ingeniería de características

La ingeniería de características busca tomar un conjunto de datos y con estos construir variables características o explicativas que se puedan emplear posteriormente en un modelo de aprendizaje de máquinas, con el fin de resolver cualquier problema de predicción o clasificación.

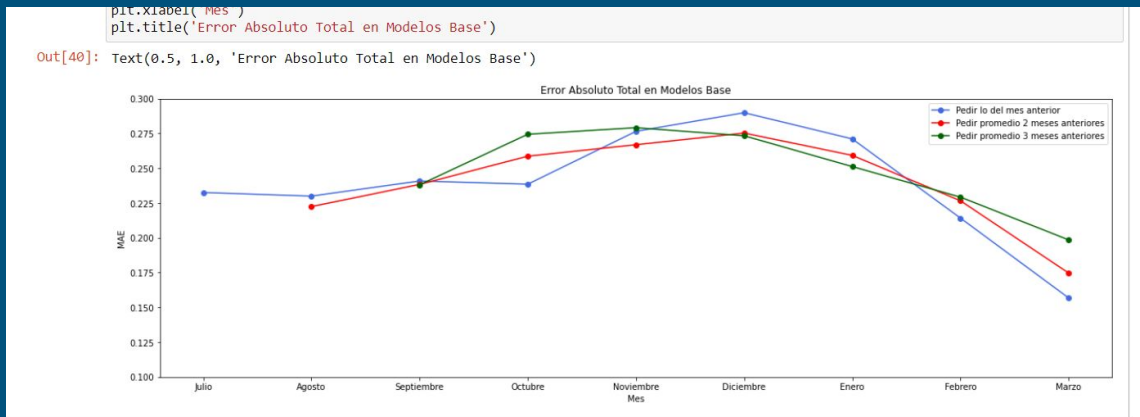
Todas estas variables, como equipo, las consideramos muy útiles para el resto del proyecto, nos guían un paso mas hacia nuestra variable de respuesta, que es la predicción de la demanda, puesto que al tener agrupados y contados estos datos con estas variables características, se vuelve información significativa y accesible para nuestras siguientes etapas.



Etapa 4: Modelado

1) Construir (codificar) el modelo de promedios móviles, cambiando 2 veces el periodo móvil.

Como se puede observar en la gráfica que obtuvimos, la línea de color rojo es la más estable por lo que es la que elegiríamos, es la más estable debido a que aunque comienza con un poco de más error no brinca tanto como la azul o la verde que suben en demasía.



PROMEDIOS MÓVILES

Los promedios móviles son promedios calculados a partir de subgrupos artificiales de observaciones consecutivas. Se utiliza cuando se quiere dar más importancia a conjuntos de datos más recientes para obtener la previsión. Cada punto de una media móvil de una serie temporal es la media aritmética de un número de puntos consecutivos de la serie, donde el número de puntos es elegido de tal manera que los efectos estacionales y / o irregulares sean eliminados.

- **Promedio móvil simple**

- Esta técnica se usa cuando se desea darle mayor importancia a un conjunto de datos recientes y de esa manera lograr obtener un pronóstico. Cada vez que se adquiere una nueva observación, esta se debe agregar al conjunto de datos y se elimina la observación que tenga más antigüedad.

- **Promedio móvil ponderado**

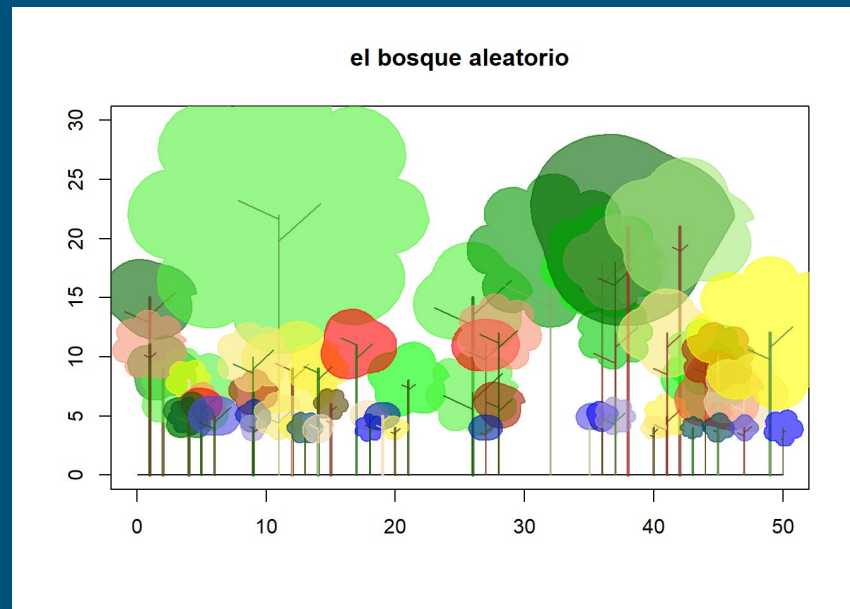
- En general se difiere que los diversos puntos de datos se pueden ponderar o asignar a un punto concreto de gran importancia. La media móvil ponderada tiene la capacidad de agregarle importancia a los puntos de datos que estén más recientes.

Día	1	2	3	4	5	6	7	8
Nivel de inventario	4310	4400	4000	3952	4011	4000	4110	4220
Promedio móvil	4310	4355	4200	3976	3981.5	4005.5	4055	4165

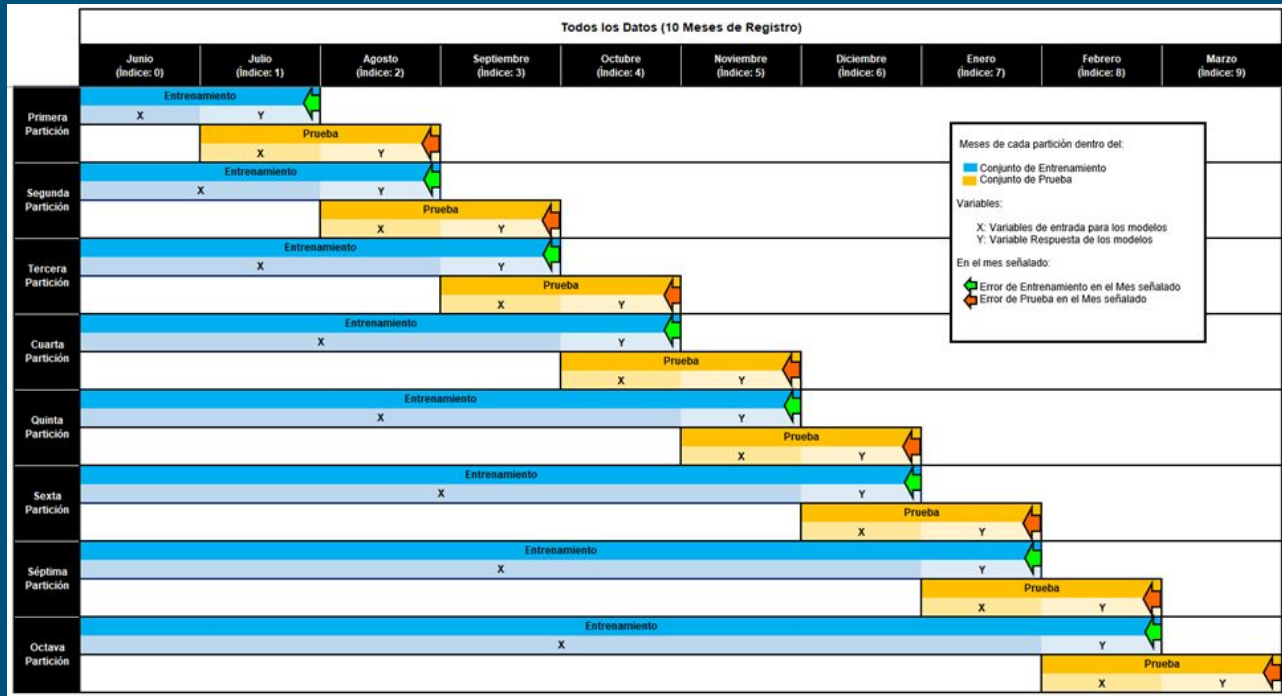
BOSQUES ALEATORIOS

CONCEPTO

Los bosques aleatorios es un algoritmo de machine learning flexible y fácil de usar que produce incluso sin ajuste de parámetros, un gran resultado la mayor parte del tiempo. Es un gran algoritmo para entrenar temprano en el proceso de desarrollo del modelo, para ver cómo se desempeña y es difícil construir un mal modelo con este algoritmo debido a su simplicidad.



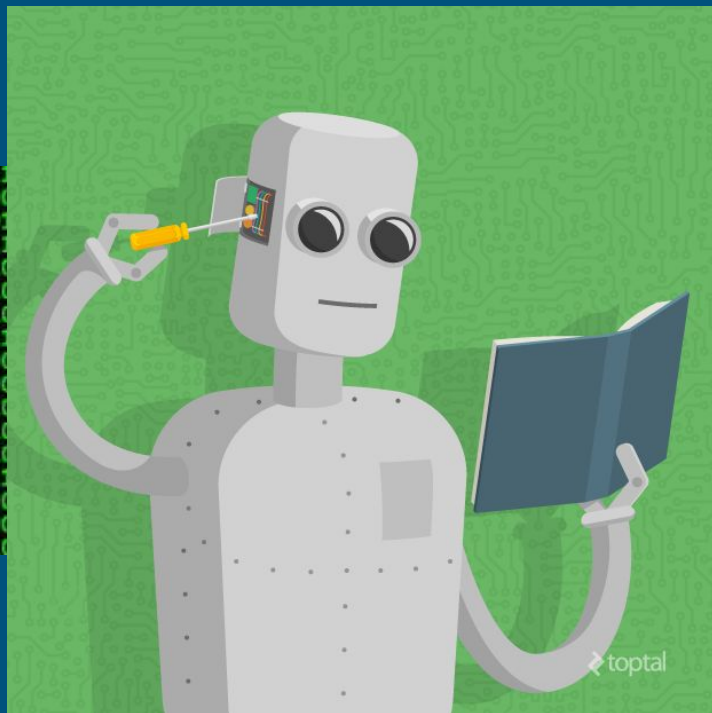
VALIDACIÓN CRUZADA PARA SERIES DE TIEMPO



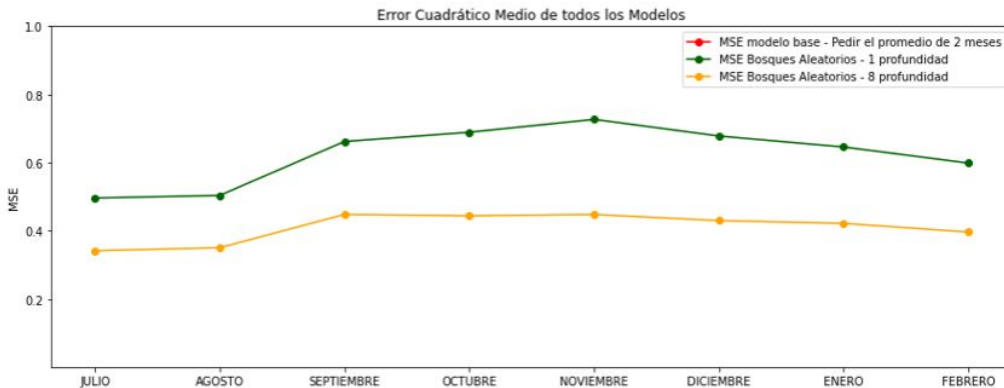
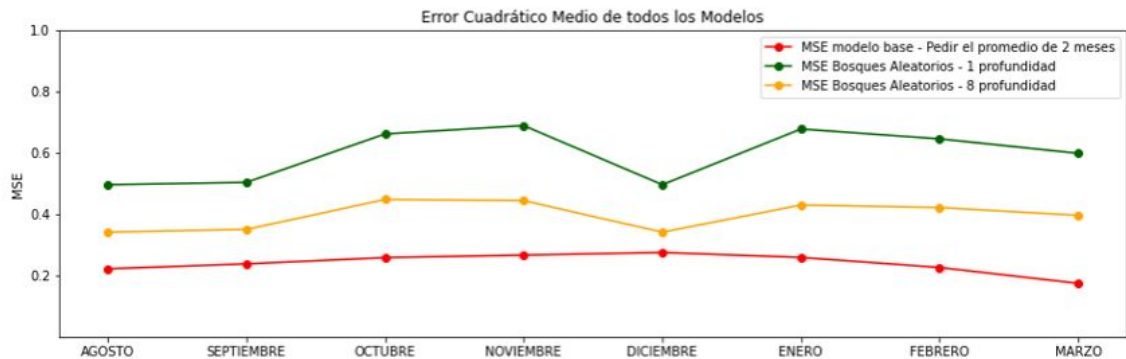
BOSQUES ALEATORIOS Y JUPYTER



```
00100000000010100011011000000100101100011
11000101110100010001111111110100000100
00101001011000011010111011010110110010001
0110110000010101100100010000111000100111
01001100101101001101100111101111011110
000110100#include <stdio.h>01101000011010
010010011000100110001010001110
10001001int main()0000101111
010101001{000011000
111001100 printf("Hello World");00001100
001000001return 42;010101110110
000110100100011000110001101000011010
0100100110111101011101110000001010001110
100010010001010110010011101110100010111
010101001110011010101110001010100011000
11100110000011011111010100111110001100
0100000111111101010010011010101110110
```



Etapa 5: Evaluación y conclusiones



Bibliografía:

Bryan Salazar López. (2019). Promedio móvil. 15/10/2020, de Ingeniería industrial online Sitio web: <https://www.ingenieriaindustrialonline.com/pronostico-de-la-demanda/promedio-movil/>

Autor Desconocido. (2019). ¿Qué es un promedio móvil?. 15/10/2020, de Minitab Sitio web: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/time-series/supporting-topics/moving-average/what-is-a-moving-average/>

Josefina Pacheco. (2019). ¿Qué es el Promedio Móvil?. 15/10/2020, de Web y Empresas Sitio web: webyempresas.com/promedio-movil/

Apple inc.. (2020). Apple. 06/09/2020, de Apple Sitio web: <https://www.apple.com/>

Edgar Sánchez. (2019). Participación de mercado de Apple y Huawei 2010-2018. 06/09/2020, de Merca2.0 Sitio web: <https://www.merca20.com/mercado-apple-huawei/>

¿PREGUNTAS?