

# VENTAS DIARIAS DE UNA EMPRESA AMERICANA

DAILY SALES OF A AMERICAN COMPANY.

Autor: Juan Camilo Varón

IS&C, Universidad Tecnológica de Pereira, Pereira, Colombia

Correos-e: [juan.varon@utp.edu.co](mailto:juan.varon@utp.edu.co)

**Resumen—** Machine Learning es una disciplina científica del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente. Aprender en este contexto quiere decir identificar patrones complejos en millones de datos. La máquina que realmente aprende es un algoritmo que revisa los datos y es capaz de predecir comportamientos futuros. Automáticamente, también en este contexto, implica que estos sistemas se mejoran de forma autónoma con el tiempo, sin intervención humana. Veamos cómo funciona.

**Palabras clave—** Machine Learning, Inteligencia Artificial, predicción, de forma autónoma.

**Abstract—** Machine Learning is a scientific discipline in the field of Artificial Intelligence that creates systems that learn automatically. Learning in this context means identifying complex patterns in millions of data. The machine that really learns is an algorithm that reviews the data and is able to predict future behavior. Automatically, also in this context, implies that these systems are improved autonomously over time, without human intervention. Let's see how it works.

**Key Word—** Machine Learning, Artificial Intelligence, prediction, autonomously.

## I. INTRODUCCIÓN

Una empresa americana tiene un comportamiento en la venta de sus productos. En el siguiente proyecto mostrare el comportamiento de estas ventas y su proyección a futuro.

## II. DESARROLLO

Inicio mi proyecto cargando un dataset que contiene información de casi 2 años de ventas diarias de productos. Los campos que contiene son fecha y la cantidad de unidades vendidas.

Utilizando Pandas, cargo y trato todos mis datos. Importo las librerías que utilizaré y realizo lectura del archivo csv.

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 %matplotlib inline
5 plt.rcParams['figure.figsize'] = (16, 9)
6 plt.style.use('fast')
7
8 from keras.models import Sequential
9 from keras.layers import Dense, Activation, Flatten
10 from sklearn.preprocessing import MinMaxScaler
11
12 df = pd.read_csv('time_series.csv', parse_dates=[0], header=None, index_col=0, squeeze=True, names=['f
13 df.head()
```

fecha

2017-01-02 236

2017-01-03 237

2017-01-04 290

2017-01-05 221

2017-01-07 128

Name: unidades, dtype: int64

Se puede notar que el dataframe que cargué con pandas tiene como índice mi primera columna con las fechas. Esto es para que me permita hacer filtrados por fecha directamente y algunas operaciones especiales.

Por ejemplo, puedo ver de qué fechas tengo datos con:

```
1 print(df.index.min())
2 print(df.index.max())
```

Los datos que obtuve de esta empresa son de ventas diarias entre los años 2017 y 2018, hasta el mes de noviembre. A continuación se visualiza cuantas muestras tengo de cada año:

```
1 print(len(df['2017']))
2 print(len(df['2018']))
```

315

289

Como esta empresa no tiene atención los domingos, se puede ver que de 2017 no se tiene 365 días como erróneamente se podría presuponer. Y en 2018 falta el último mes, que será lo que trataré de pronosticar.

### VISUALIZACIÓN DE DATOS

Estos datos me puede brindar pandas con describe():

```
1 df.describe()
```

```
count 604.000000
mean 215.935430
std 75.050304
min 51.000000
25% 171.000000
50% 214.000000
75% 261.250000
max 591.000000
Name: unidades, dtype: float64
```

Son un total de 604 registros, la media de venta de unidades es de 215 y un desvío de 75, es decir que por lo general estaré entre 140 y 290 unidades.

De hecho, se puede aprovechar el tener índice de fechas con pandas y sacar los promedios mensuales:

```
1 meses = df.resample('M').mean()
2 meses
```

fecha

2017-01-31 203.923077

2017-02-28 184.666667

2017-03-31 182.964286

2017-04-30 198.960000

2017-05-31 201.185185

2017-06-30 209.518519

2017-07-31 278.923077

2017-08-31 316.000000

2017-09-30 222.925926

2017-10-31 207.851852

2017-11-30 185.925926

2017-12-31 213.200000

2018-01-31 201.384615

2018-02-28 190.625000

2018-03-31 174.846154

2018-04-30 186.000000

2018-05-31 190.666667

2018-06-30 196.037037

2018-07-31 289.500000

2018-08-31 309.038462

2018-09-30 230.518519

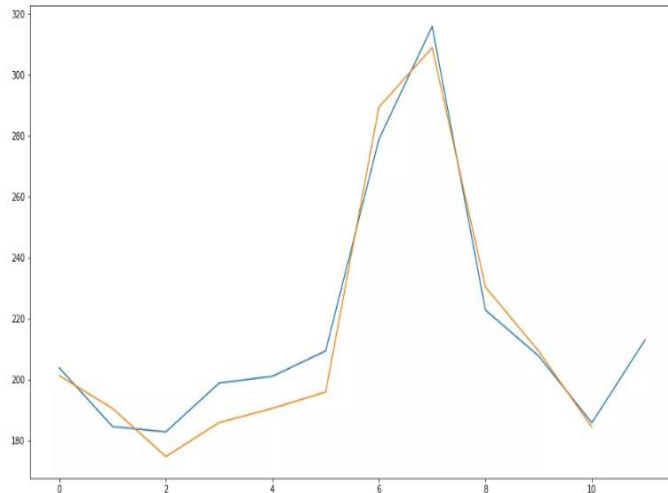
2018-10-31 209.444444

2018-11-30 184.481481

Freq: M, Name: unidades, dtype: float64

Visualizando esas medias mensuales:

```
1 plt.plot(meses['2017'].values)
2 plt.plot(meses['2018'].values)
```



También se puede ver que en el año 2018 (naranja), la línea se comporta prácticamente igual. Es decir, que pareciera que se tiene una estacionalidad. Por ejemplo podría llegar a pronosticar que “el verano de 2019 también tendrá un pico de ventas”.

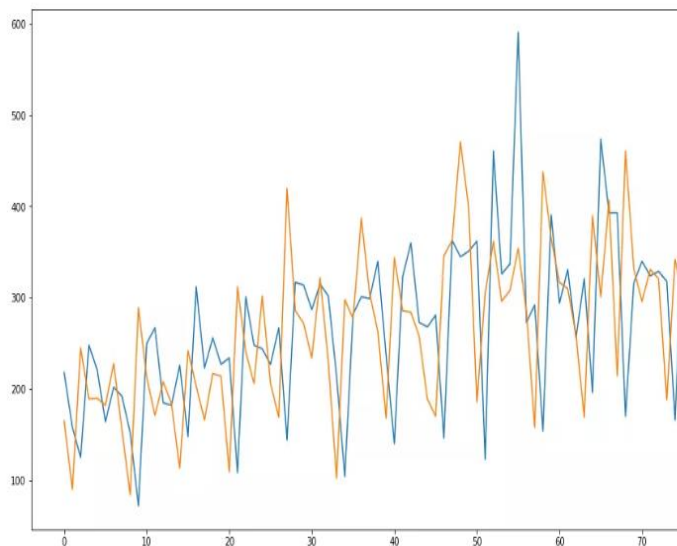
### REFERENCIAS EN LA WEB

<https://www.apd.es/que-es-machine-learning/#:~:text=Machine%20Learning%20o%20Aprendizaje%20autom%C3%A1tico,de%20datos%20en%20su%20sistema.>

<https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico>

En la siguiente gráfica se pueden ver las ventas diarias (en unidades) entre junio y julio:

```
1 verano2017 = df['2017-06-01':'2017-09-01']
2 plt.plot(verano2017.values)
3 verano2018 = df['2018-06-01':'2018-09-01']
4 plt.plot(verano2018.values)
```



### CONCLUSIONES

Se puede ver la línea azul (año 2017), se tiene un inicio de año con un descenso en la cantidad de unidades, luego comienza a subir hasta la llegada del verano americano en donde en los meses junio y julio tenemos la mayor cantidad de ventas.

Finalmente vuelve a disminuir y tiene un pequeño pico en diciembre, con la llegada de la Navidad.