

**UNIDAD EDUCATIVA TUMBACO**

FORTALECIMIENTO DE GESTIÓN INSTITUCIONAL

FGI

**MATEMATICAS**

**PROYECTO DE EXPLORACION**

Análisis del papel que tiene las matemáticas en la inteligencia artificial al momento de interpretar datos

**2100656665**

Jorge Andrés Ortega Garrido

2266 palabras/15 hojas

**2021 - 2022**

**EVALUACIÓN**

|  |
| --- |
| **Comentario**  Al realizar esta investigación me sentí muy comprometido y feliz de realizarla, pues intente meter la programación para realizar algunas cosas y ver como miles de datos se pueden simplificar en pocos números. Por otro lado si se acomplejo un poco llevar a cabo esta investigación puesto que había que investigar varias cosas para abarcarla de la mejor manera y sobre todo que sea entendible a un nivel académico |

|  |  |
| --- | --- |
| Calificación: | Docente responsable:  Nombre:  Firma: |

La inteligencia artificial (IA) ha tenido más renombre en los últimos años pues es capaz de hacer cosas increíbles casi hasta el punto de remplazar a un humano, tiene varias aplicaciones en la vida cotidiana y también ha formado un papel importante en nuevos descubrimientos científicos como descubrir nuevos materiales, interpretar la primera imagen de un agujero negro y de una nueva revolución que promete la conducción autónoma de los automóviles, pero… ¿cómo es que posible que la IA se capaz de lograr algo así de complejo? Como ya lo dijo (Madiha Jamal, 2020) “La IA no es magia; son solo matemáticas.” Siendo esto un claro ejemplo que las matemáticas están en todos lados. A través de la matemática, en especial: el álgebra lineal, la probabilidad y el cálculo junto con informática se forman para forjan las bases de lo que es la inteligencia artificial: pero esto no solo queda aquí… se necesitan datos que llenen de conocimiento a esta inteligencia para que sea capaz de arrojar resultados y pensar de manera autónoma siendo esto cada vez más precisos, a este proceso en el cual la inteligencia ya es capaz de pensar y hasta cierto punto razonar se lo denomina “machine learning” que literalmente significa “aprendizaje autónomo”. Los datos que se introduzcan al Machine Learning varían de acuerdo al propósito que tenga.

La inteligencia artificial fue creada con el propósito de pensar y razonar como un humano real, y con ello su construcción y conceptos toma parte de como funciona el cerebro para lograr una inteligencia de tal magnitud. Las redes neuronales son como las neuronas que el cerebro tiene y cumplen con el objetivo de comunicar e interpretar datos de manera eficaz y rápida al haber bastantes conexiones neuronales la capacidad que tienen de transmitir información se hace muy rápida siendo esta una característica primordial para la creación y funcionamiento de la inteligencia artificial. Para llegar a comprender mejor el machine learning, cuando recién está en desarrollo es como un niño al que se le debe llenar de conocimiento para que pueda hacer algo relacionado con

el tipo de comportamiento y educación que recibió el niño, este basara sus acciones futuras en base a los conocimientos que se le ha dado. Asimismo, funciona la IA, si se desea que haga cálculos matemáticos complejos, será necesario alimentarla de conocimiento en forma de datos para que sea capaz de comprenderlos y posteriormente dar un resultado preciso.

En esta investigación hablare del índice de depresión a través de los factores que motivan la aparición de depresión, tales como: edad, nivel de educación, gastos de manutención y salario. Se usarán basada en antiguos datos del el 2014 que proporcionarán un conocimiento adecuado para el machine learning.

Para desarrollar el machine learning se usará un lenguaje de programación bastante conocido dentro de esta rama, Python se usará para crear el “aprendizaje autónomo” a través de este se crearán las redes neuronales donde se alojaron los datos antes mencionados. Los datos tendrán una estructura bastante detallada que seguir, con 1400 encuestados que respondieron de acuerdo a la situación en la que se encuentran Todos estos datos serán en forma de números que se interpretarán. Se hará uso de módulos que ayuden a armar el machine learning de esta IA, los módulos serán “pandas” , “numpy” y “seaborn” que brindarán el análisis de datos, y “matplotlib” que ayudará a entender estos datos de manera gráfica, dichos módulos de Python brindarán ayuda al crear las redes neuronales necesarias para obtener un resultado preciso.

El proyecto contemplará las siguientes acciones:

* Recoger datos en fuentes verídicas sobre un lugar determinado, estos datos ejemplificarán la situación de una persona teniendo en cuenta mas de 10 factores para un análisis adecuado.
* Análisis breve de los datos.- examinar los datos recogido y ver de que tipo son para armar un esquema del algoritmo que debe tener la IA
* Crear la IA con los parámetros y factores dados
* Introducir datos en la IA
* Analizar los datos obtenido y ponerlos en “matplotlib” para convertir los datos en gráficos precisos.

A través de estos pasos se obtendrán conclusiones en relación al problema planteado en esta investigación

**INFORMACION Y MEDICIONES**

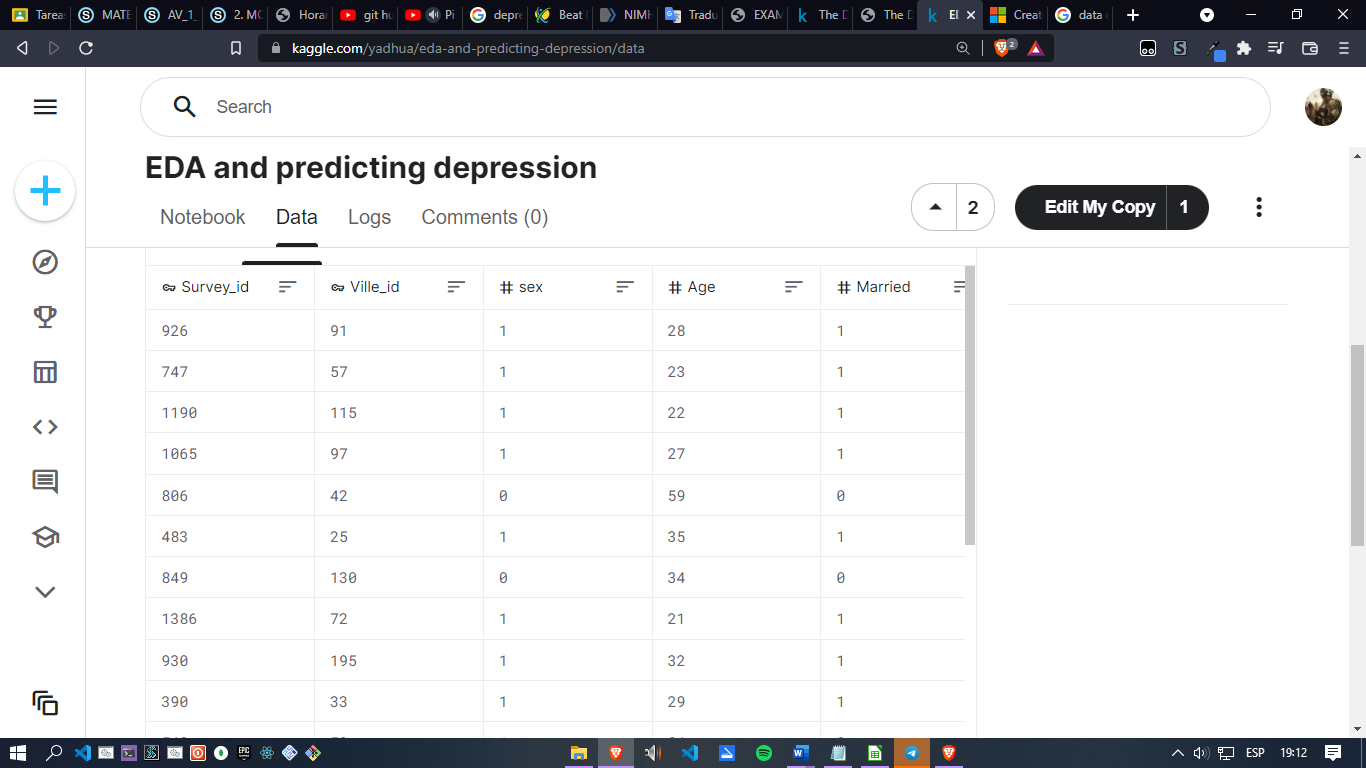
Para poner en marcha esta investigación se hará el uso de una base de datos proporcionada por Kaggle.com. Está contiene información valiosa acerca de datos estadísticos, en este estos datos serán acerca de los factores que fomentan la depresión, se encontraron varias bases de datos que tienen diferentes parámetros como: Sexo, edad, casado, nivel de educación, total de miembros de la familia, activo duradero activo salvado, gastos de manutención, salario, poseedor de negocio y estado en el que se encuentra, deprimido o no deprimido.

Tabla 1. Vista general de la forma en la cual los datos están organizados

Fuente: <https://www.kaggle.com/yadhua/eda-and-predicting-depression/data>

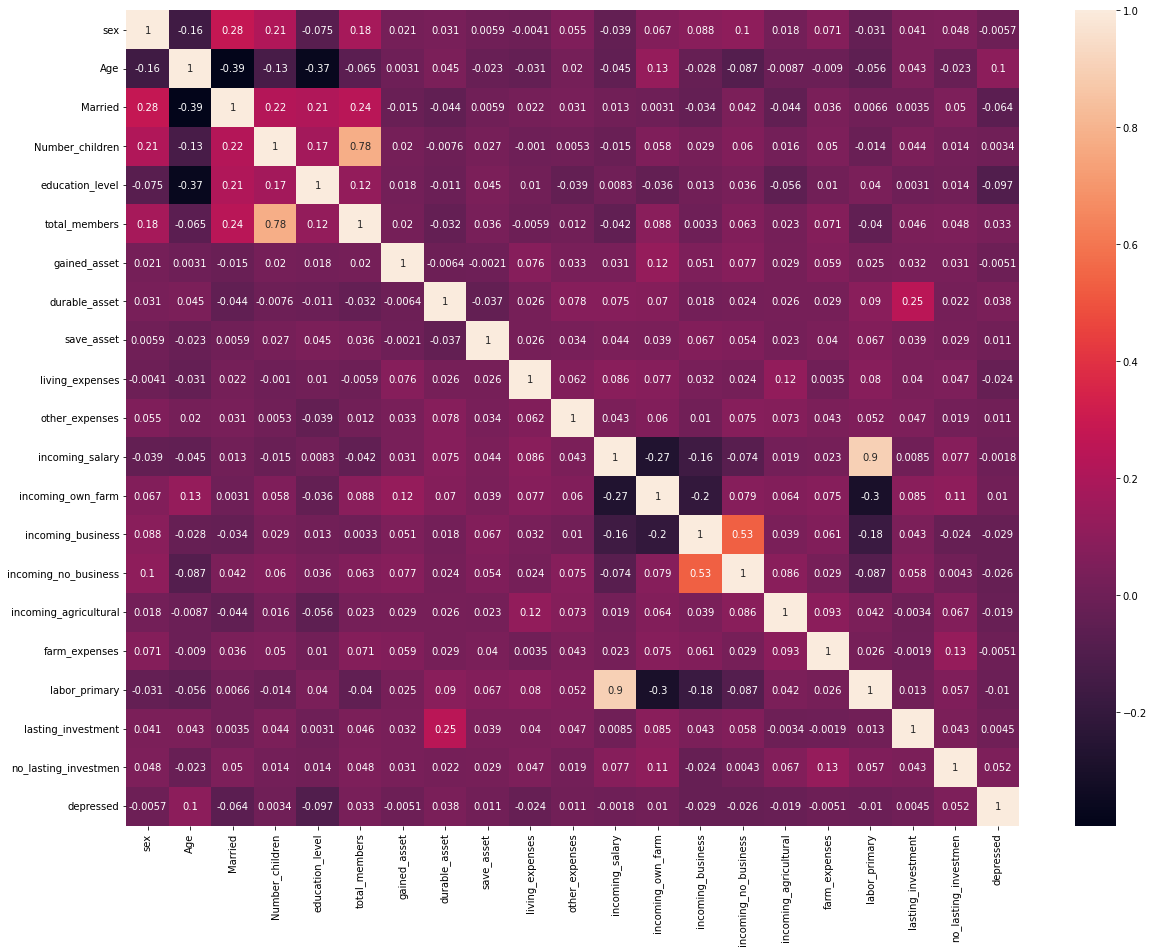


Imagen que toma todos los factores y parámetro reunidos en un numero total basado en las 1429 personas encuestadas.

Fuente: <https://acortar.link/jgDagt>

Las filas de datos contienen un total de 20 factores que serán analizados. Para algunos datos, como género y la respuesta a la pregunta de si el encuestado se siente o no deprimido, se usa el valor del 0 y 1, el 0 representa un valor de no y el 1de sí. En cuanto al genero se usa el 0 para representar genero femenino y 1 para el masculino.

A través de la pregunta de si está o no deprimido se sentarán la importancia de los factores que fomenta que el encuestado haya dicho que si está deprimido. Por ejemplo, si existe un mayor índice de las persona que se sienten deprimida a partir de cierta edad, la IA detectará este patrón y arrojara resultados en base a la importancia y similitud entre los datos.

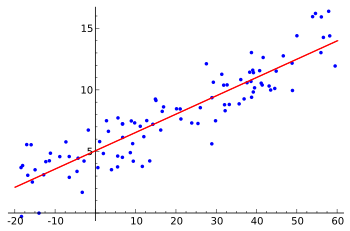
|  |  |
| --- | --- |
| Factor | Importancia |
| Edad | 0.115239 |
| Nivel de educación | 0.102920 |
| Gastos de manutención | 0.078137 |

Tabla que muestra los 3 factores que mas afectan sobre la sensación de sentirse deprimido. Estas 3 variables se analizan a partir de los 1429 participantes y la relación sobre la depresión.

Para obtener un resultado más preciso se usará el módulo “sklearn” de Python, este brindará la posibilidad de encontrar un patrón entre los datos estadísticos, Por ejemplo: la media de la edad las personas que se sienten deprimidos y las que no, la tasa de probabilidad, que entre mas o menos hijos tenga una persona, varie la sensación de sentirse deprimido ya sea que bajen o aumenten las posibilidades, y el resto de variables que se encuentran en la tabla se manejaran de manera global para relacionarlos y ejemplificar cuales variables inducen a una persona a sentirse deprimida, a través del análisis de la importancia los factores se clasificaran de manera ascendente, entre mayor sea el numero de la variable que representa la importancia, mayor será el impacto que tendrá sobre la depresión.

**Procesos matemáticos**

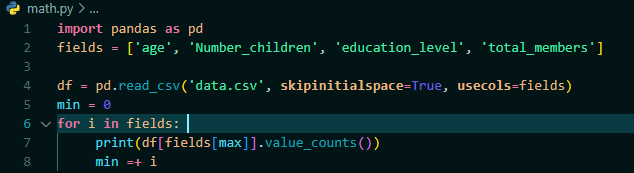
Para cumplir con la presente investigación, se hará uso procesos matemáticos para refutar con datos verídicos la información a tratar.

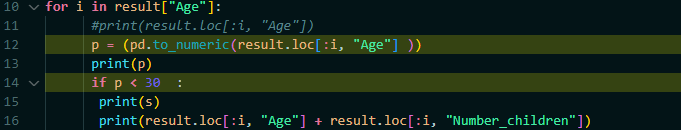
**Regresión lineal**

Modelar los datos a través de variables para encontrar una relación. Será otro modelo que se usará para medir el comportamiento de las variables y contrastarlos con el modelo de la IA. La variable independiente será la edad y las variables dependientes serán estos 5 factores: Sexo, nivel de educación,

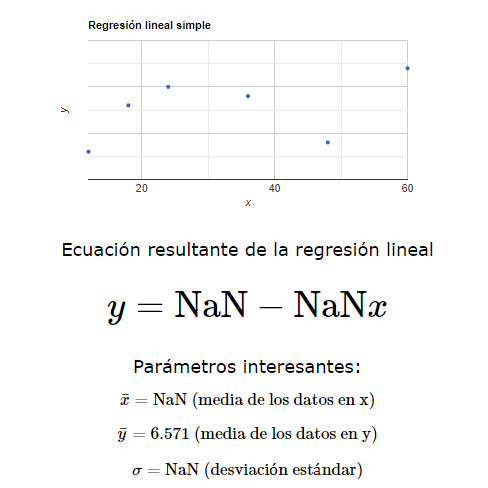
gastos de manutención, total de miembros de la familia.

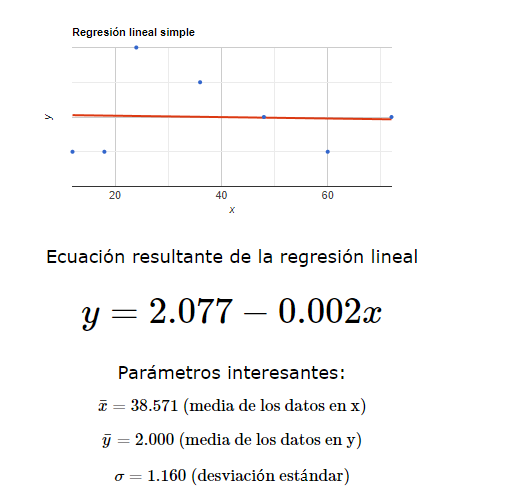
Para agrupar los datos y colocarlos sobre el modelo de regresión lineal, se usará Python.

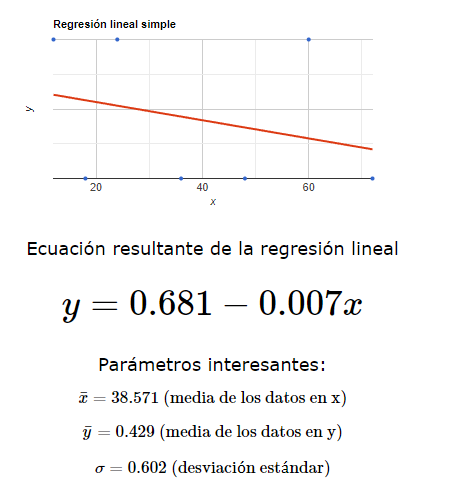


Para relacionar de manera directa las variables dependientes sobre la variable dependiente, se empleara este fragmento de código: 

Donde en la línea 16, la segunda palabra entre comillas se cambiará para conocer exactamente la correlación entre la edad y las demás variables.

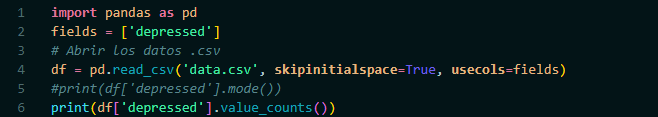
A continuación, se muestra la tabla con valores en base a la **edad** y el **nivel de educación**, dando como conclusión que las personas con una mayor edad tienden a tener un mayor nivel educativo que las personas con menor edad.

Ahora, se usa como variable independiente, la **cantidad de hijos** que una persona tiene en base a la edad.

En la gráfica, se aprecia que a mayor edad la cantidad de hijos irá disminuyendo, otra relación que se encuentra es que entre la edad de 20-30 la cantidad de miembros familiares llega a aumentar.

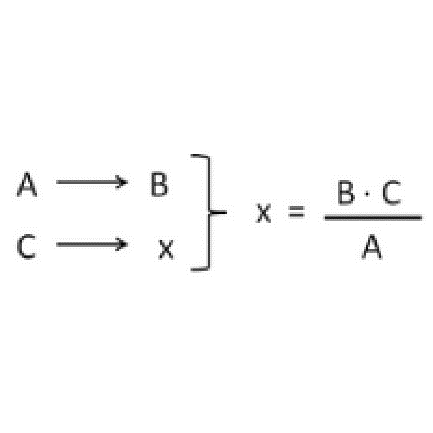
En la siguiente gráfica se muestra la relación entre la edad y el género.

Con esto datos obtenidos, se procederá a establecer cual es el factor que fomenta, la depresión en las personas.



Código que divide la respuesta de si o no a la pregunta “te sientes deprimido”



Total, de respuestas: 1429.

Se uso la **regla de 3**

1429 100%

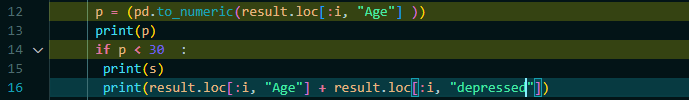
1191 ?

= 83,34%

1191 representa el **83,34%**.

238 representa el **16,66%** (resultado de 100 – 83,34)

Una vez obtenida los datos, se relacionará de manera directa el principal los factores que contribuyen a la sensación de sentirse deprimido.



El código base, filtrará los datos por edad, si la persona tiene mas de 30 años recolectará si se siente deprimida o no.

Se trabajará solo con las 238 personas que han respondido que si se sienten deprimidas.

En base a la edad, se filtraron los datos, gracias a la herramienta Keggle.com se pasaron los datos por Machine learning que arrojo resultados en base a la importancia y tendencia de personas que se sienten deprimidas y cual factor es el mas común que comparten. El numero 1 representa una importancia máxima, mientras que el 0, una nula por dicho motivo los numero van acompañados de decimales para establecer dicha importancia.

**Cálculo del coeficiente de correlación**

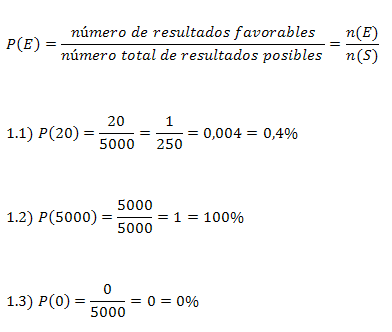
n=cantidad de datos

x=variables dentro de la depresión (ej: edad, sexo, educación)

y=importancia de variables dentro de la depresión (ej: edad: 0.11, educación:0.07 )

La correlación de los datos es positiva y moderada, ya que se considera la formula y el resultado del coeficiente de correlación que cumple la siguiente condición . Con este valor se deduce que la importancia neta de todas las 3 variables (sexo, educación y cantidad de hijos) infiere con un 0.71

**Probabilidad**

 Calcular la posibilidad entre todas las variables; sobre que estas influyan sobre la depresión y cuál es la posibilidad que a partir de cierta edad o circunstancia la depresión tenga más posibilidades de aparecer o tener un efecto preocupante.

Por otra parte la inteligencia artificial, también puede ayudar a predecir sentimientos humanos de manera globalizada, en este caso será sobre cuales factores son los que más se repiten cuando una persona se siente deprimida.

Para las personas con una edad inferior a 30 que posee mas de 2 hijos la posibilidad de que se sienta deprimido es de: 45%

= 0.45 %

Se tomó como base, que de cada 100 encuestados menores a 30, 45 de ellos contestaron que se sienten deprimido, cave recalcar que este no es la única variable, existen 4 mas que hay que tener en cuenta.

Personas con edad superior a 30 años, la probabilidad que padezca deprimido por tener hijos es del: 18%

= 0.18 %.

Si se continua asi sucesivamente obtenemos el resultado de que el mayor factor que afecta a una persona para sentirse deprimida, es el nivel de educación seguido de la cantidad de hijos. Todo ello muestra una tasa de diferencia en base a la edad, pues si una persona mayor (>60) tiene mas familiares e hijos la sensación de sentirse deprimida no estará allí. Por el contrario, si una persona joven o adulta posee una gran cantidad de familiares es una factor mas que influirá sobre la sensación de sentirse deprimido

# Validez

Para esta investigación se usaron los procedimientos pertinentes para encontrar la fórmula que permita predicciones de valores futuros, los datos provienen de una página confiable en donde se encuentra información muy variada sobre distintos datos, el fin último de la página es proporcionar información para que programadores sea capaces de crear modelos de machine learning, lo que otorga una confiabilidad en los datos, por otro lado Kaggle es una subsidiaria de la empresa Google SA, lo que da confianza en la validez de la información. Al mismo tiempo al usar herramientas de programación se puede abarcar una gran cantidad de datos que permitan ser precisos, con mas de 1400 encuestados se logra una precisión notable sobre las variables que fomentan la depresión en una persona, junto con esto al pasarlo por un esquema de inteligencia artificial llegan ha ser datos bastante apegados a la realidad.

**Bibliografías**

Mathurin Aché. (2020). World Happiness Report up to 2020. Retrieved January 22, 2021, from Kaggle.com website: https://www.kaggle.com/mathurinache/world-happiness-report/tasks?taskId=1259

‌ Anaconda | Individual Edition. (2021). Retrieved January 22, 2021, from Anaconda website: https://www.anaconda.com/products/individual

‌ J. Gabriel Molina / María F. Rodrigo. (2009). *T. 9 – El modelo de regresión lineal*. Retrieved from <http://ocw.uv.es/ciencias-de-la-salud/pruebas-1/1-3/t_09nuevo.pdf>

Welcome to Python.org. (2021, March). Retrieved March 11, 2021, from Python.org website: <https://www.python.org/>

Tipos de funciones. (2015, September 4). Retrieved March 11, 2021, from Universo Formulas website: https://www.universoformulas.com/matematicas/analisis/tipos-funciones/

‌ wikiHow. (2010, November 16). Cómo calcular un incremento en porcentaje. Retrieved March 11, 2021, from wikiHow website: https://es.wikihow.com/calcular-un-incremento-en-porcentaje

Chakraborty Ananya (1 de junio de 2021) ¿Cómo aprender matemáticas para el aprendizaje automático? ¿Qué conceptos necesitas dominar en ciencia de datos?, recuperado de: <https://www.analyticsvidhya.com/blog/2021/06/how-to-learn-mathematics-for-machine-learning-what-concepts-do-you-need-to-master-in-data-science/>

W3Schools (SN/F) Machine Learning - Mean Median Mode, recuperado de: <https://www.w3schools.com/python/python_ml_mean_median_mode.asp>