

PROYECTO DE GRADO

APLICACIÓN DEL MACHINE LEARNING Y BIG DATA EN EL ANALISIS DE LA  
HUELLA DE ASFALTO EN CULTIVOS DE MAÍZ

UNIVERSIDAD CATOLICA DE MANIZALES.  
FACULTAD DE INGENIERIA Y ARQUITECTURA.  
INGENIERIA DE TELECOMUNICACIONES.

JHON EISENHOWER SABOGAL GARCÍA.

2021.

## Contenido

<b>OBJETIVOS</b> .....	<b>3</b>
Objetivo General .....	<b>3</b>
Objetivo Específicos .....	<b>3</b>
<b>JUSTIFICACIÓN</b> .....	<b>3</b>
<b>DESCRIPCIÓN DEL PROBLEMA</b> .....	<b>4</b>
<b>MARCO REFERENCIAL</b> .....	<b>5</b>
Factores ambientales .....	<b>5</b>
Phyllachora maydis y fitopatología.....	<b>7</b>
Machine learning .....	<b>9</b>
<b>Big Data</b> .....	<b>12</b>
Aplicaciones del Big Data .....	<b>12</b>
Redes Neuronales artificiales –RNA .....	<b>14</b>
<b>APLICACIÓN DEL MACHINE LEARNING Y BIG DATA EN EL ANALISIS DE LA HUELLA DE ASFALTO EN CULTIVOS DE MAÍZ</b> .....	<b>16</b>
DATASET .....	<b>16</b>
<b>DESCRIPCION DEL ALGORITMO Y PROGRAMACION</b> .....	<b>17</b>
PERCEPTRON MULTICAPA (MLP).....	<b>18</b>
REGRESOR DE PERCEPTRON MULTICAPA .....	<b>18</b>
Por qué Python? .....	<b>20</b>
CODIGO DE PROGRAMACION PYTHON .....	<b>21</b>
<b>DESCRIPCION DE HALLAZGOS Y CONCLUSIONES</b> .....	<b>23</b>
CONCLUSION DE LA CORRELACION: .....	<b>26</b>
CONCLUSIONES DE LA RED NEURONAL .....	<b>29</b>
<b>TABLA DE REFERENCIA DE IMÁGENES</b> .....	<b>30</b>
<b>Referencias</b> .....	<b>30</b>

## **OBJETIVOS**

### **Objetivo General**

Mostrar las aplicaciones del Machine Learning y del Big Data en el estudio de cultivos por medio de Python.

### **Objetivo Específicos**

- Identificar si hay una relación que se encuentra entre los factores ambientales y el aumento de la Phyllachora Maydis en una plantación de maíz.
- Establecer la importancia que tiene la ciencia de los datos en los diferentes ámbitos de la vida cotidiana.
- Conocer sobre las redes neuronales y su importancia en la predicción de eventos.

## **JUSTIFICACIÓN**

El presente trabajo de grado tiene como la finalidad de investigar la relación que se encuentra entre los factores ambientales (abióticos) en el crecimiento de patógenos en plantaciones, como la variación de estos factores pueden afectar los índices de severidad en los cultivos, en nuestro proyecto tomamos como especie muestral el maíz, esta es un planta muy importante es nuestra cultura americana y es considerada uno de los alimentos esenciales para la supervivencia humana junto a otros cereales como el trigo, se cree que fue originario en Mesoamérica pero se expandió en los tiempos de las colonizaciones británicas y a españolas a el resto el mundo, prescindir de un alimento como el maíz podría traer graves consecuencias en la alimentación de muchos países pobres y subdesarrollados, donde es una de los alimentos principales de las canastas alimenticias básicas, entonces aquí es donde se hace esencial investigaciones y desarrollos que permitan la toma de decisiones anticipadas que favorezcan la mejora de la producción del maíz o cualquier otro cultivo. Es importante comenzar a implantar técnicas que mejoren la salud de las plantaciones previniendo condiciones que generen baja calidad en los granos y a un futuro volver a plantar la especie en la misma tierra donde se hizo la siembra. El maíz es una de los alimentos que exportamos al mundo en Colombia genera miles de empleos y fuente de sustento a nuestros agricultores, nuestro país es un país en vía de desarrollo que tiene como fortaleza la exportación de cultivos como el café o el mismo maíz, comenzar a normalizar tecnologías que faciliten el proceso de siembra y cuidado de las plantas alimenticias generaran desarrollo en la región, el estado colombiano indica que busca que Colombia sea la próxima Silicón Valley en Sudamérica (Forbes, 2021), por ello es indispensable pensar en soluciones tecnológicas que optimicen los procesos más fuertes que tenemos en nuestra económica como lo es la agricultura, tener predicciones y anticiparnos a eventos es una situación ideal

debido a que podremos controlar los costes en compuestos químicos para el control de plagas, la contaminación de recursos hídricos y la erosión de los suelos haciendo que nuestras tierras pierdan sus componentes, es una gran apuesta en el presente y futuro para nuestra nación, la tecnología aplicada será una fuente de desarrollo debido a que mejoraremos nuestros procesos haciéndolos más baratos, más productivos y escalables.

El Machine Learning y el Big Data en la agricultura colombiana será una gran herramienta que facilitara la vida de nuestros campesinos, la posibilidad de predecir comportamientos de cultivos, de patógenos y climáticos harán que podamos hacer mejor la siembras de especies como el maíz, llevar un seguimiento anual de los datos que nos entrega nuestras plantaciones nos hará tomar mejores decisiones, cuidando de mejor forma el medio ambiente y la calidad de las producciones, traerá mayores inversores a nuestros productos, generara una mayor cantidad de empleo y sostenibilidad de las familias, es indispensable generar un cambio en nuestro país pero aplicando nuestros conocimientos en beneficio de la sociedad. (Frattini, 2019)

## **DESCRIPCIÓN DEL PROBLEMA**

En Colombia tenemos como principal fortaleza o fuente de ingresos la agricultura, son las actividades económicas que mueven el PIB del país, somos conocidos como el país cafetero precisamente por nuestra gran fortaleza en las técnicas y cultivos de los mejores cafés en el mundo, además de las grandes fortalezas en la exportación de otros productos agrícolas nativos de nuestras tierras como son los bananos y los plátanos, las flores, además de otros alimentos que se cultivan en nuestro país en Colombia en el año 2020 se tuvieron ingresos al PIB nacional por \$7.340 millones (Semana, 2020). Sin embargo podríamos poder tener mayor crecimiento en las exportaciones a otros países si fortaleciéramos los procesos y el apoyo al campesino, el cultivador en Colombia no solo se debe enfrentar a los problemas sociales que ha tenido históricamente nuestro país, igualmente debido a las diferentes condiciones climáticas que hay en cada región y de la composición de los suelos, ha llevado que tanto como se presenta mayor variedad de especies a cultivar también se presente mayores patógenos que afecten el crecimiento de las plantas, en Colombia tenemos todas las condiciones climáticas dado la presencia de los pisos térmicos, esto favorece la gran diversidad de enfermedades que afectan los cultivos de nuestros agricultores, esto desemboca en menor calidad y producción de los cultivos como es el caso de la especie que va ser estudiada en este proyecto como es el Maíz, el maíz desde los tiempo precolombinos ha sido la fuente de alimento y de sustento económicos de muchos habitantes de nuestro país que han dirigido su actividad económica a cultivar este importante especie. Debido a las condiciones climáticas donde se siembra el maíz son favorables en nuestra región (de 8°C hasta los 30°C) (Infoagro, s.f.), también lo es para la mancha de asfalto (*Phyllachora Maydis*) uno de los principales

patógenos que causan bajas producciones en los cultivos debido a que afecta la calidad de los granos de maíz y causa la muerte también de las plantas, los tratamientos para estas enfermedades son muy difíciles cuando ya la plantación tiene una alta presencia de este hongo, son muy costosos e igualmente ineficaces, los insumos químicos (fungicidas) pueden ser un gasto que no puede ser muchas veces realizado por los cultivadores más pobres entonces estos puede terminar en la pérdida total de su cultivo y sustento económico.

Idealmente la mejor forma de tratar esta enfermedad es la prevención de la misma, sin embargo la falta de orientación a los agricultores de las herramientas que puede utilizar, como también de estudios especializados de los suelos hacen que los cultivos en Colombia todavía sean arcaicos, la optimización tecnológico puede generarles grandes progresos al agro colombiano sin embargo es necesario del apoyo gubernamental y de inversiones privadas, tenemos gran potencial agrícola sin embargo es indispensable comenzar a aplicar el conocimiento y avances tecnológicos para generar más oportunidades a nuestra gente además de corregir los problemas causante de prácticas culturales que afectan el ecosistema como es la quema de los cultivos cuando termina la cosecha buscando que en la próxima siembra no hayan malezas y mejore la calidad del producto a corto plazo, sin embargo a largo plazo mata los nutrientes de la tierra y genera altas emisiones de carbono que contaminan nuestro planeta. (Climate & CAC, 2015)

## **MARCO REFERENCIAL**

### **Factores ambientales**

Los Factores ambientales son todos los parámetros que interactúan entre si y condicionan el equilibrio de los organismos vivos en nuestro planeta, estos factores se clasifican en dos tipos los cuales son los Bióticos y los Abióticos. Los Factores Bióticos son los factores dados por los mismos seres vivos y la interacción que tienen entre ellos, están también los factores Abióticos que son los dados por los elementos que no tienen vida pero que en su misma dinámica afectan el desarrollo de la vida que hay en nuestro planeta. (Rodriguez, 2020)

#### **Factores Bióticos**

En este tipo de factor tenemos tres clases en las cuales se clasifican así:

##### **-Productores**

Son aquellos organismos que generan materia orgánica a partir de la transformación de materia inorgánica ejemplo las plantas que realizan fotosíntesis.

#### -Consumidores

Son aquellos que se alimentan de los organismos productores o se alimentan de los mismos consumidores un ejemplo son los animales de pastoreo y del segundo tipo un lobo que se alimenta de otros animales.

#### -Descomponedores

Son aquellos que se alimentan de materia orgánica en descomposición generando al mismo tiempo nutriente.

#### Factores Abióticos

En este tipo de factores que son el eje del estudio de este proyecto tenemos 2 principales tipos del tipo físico o químico.

#### Factores químicos

##### -Agua

Es un compuesto fundamental para todos los seres vivos, casi todos los seres lo tienen y es una sustancia fundamental para la vida.

##### -Suelo

El suelo contiene miles de minerales esenciales para la vida de los organismos.

##### -Aire

El aire es indispensable para la respiración de muchos seres vivos y para la fotosíntesis de la planta dado sus componentes ricos en oxígeno y dióxido de carbono.

##### -Sol

Es una de las piedras fundamentales para la vida de los organismos, el sol aporta la energía para el desarrollo de la gran mayoría de seres.

#### Factores Físicos

##### -Temperatura

La temperatura ayuda en la dinámica de adaptación de los diferentes seres vivos en el planeta, dependiendo de cada condición permite la vida de diferentes especies.

##### -Presión Atmosférica

Esencial para establecer la cantidad de oxígeno que se encuentra en el agua y para que los organismos internamente se adapten a los espacios donde se encuentren habitando.

##### -Lluvia y Humedad

La lluvia permite el flujo del agua en el planeta además de erosionar la tierra para que absorba los minerales presentes en el ecosistema.

## Phyllachora maydis y fitopatología



[1]

La mancha de asfalto es una enfermedad que afecta de gravedad los cultivos de maíz y se genera principalmente por la presencia del Hongo *Phyllachora Maydis*, el cual es un patógeno que puede causar la muerte de la planta generando problemas muy importantes en las producciones de los cultivos de maíz, los principales síntomas que puede tener una planta con una infección fúngica de este patógeno es que comienza a crecerle puntos negros en las hojas, poco a poco va esparciéndose por la superficie de la planta generando halos de color amarillo los cuales causan una gran sequedad debido que este hongo se alimenta de los azúcares que la planta hasta conduciendo a la muerte de la planta. (Pereyda-Hernández, 2008)

La mancha de asfalto tiende a favorecer su aparición en temperatura que oscilan desde los 17 grados a 22 grados promedio, la humedad debe ser alta en el lugar donde se encuentre la plantación además de que se conoce de que suele presentarse en regiones donde la altura sea entre 1300 a 2200 metros además de que los suelos donde se realice el cultivo del maíz sean muy nitrogenados.

Es una enfermedad muy crítica para los cultivadores del maíz en el mundo debido que la presencia de esta infección en sus cultivos puede significar pérdidas que van desde el 30% al 100% de la producción, además del tratamiento es bastante costoso debido a los fungicidas que se deben usar, además del monitoreo que se debe revisar debe ser exhaustivo para evitar de que pase se vea afectada toda la plantación. (R, 2015)

### Data Science

La ciencia de datos es una ciencia que no ha podido tener una definición consensuada por parte del mundo científico, matemático e Ingenieril, la aproximación más cercana a una definición que se le puede dar a la Data Science es que es la ciencia que tiene la finalidad de a partir de set de datos obtener conocimiento o aprendizaje esto con la finalidad de mejorar la toma de decisiones, todo ello ha sido aplicado de forma continua por muchas empresas por ejemplo para crear productos que se ajusten a las necesidades de las personas, en empresas de ventas electrónicas para generar publicidad que llame la atención de los clientes que navegan por sus plataformas, hasta en las campañas políticas para encontrar propuestas que satisfagan o atraigan a las mayorías. (Lages, 2018)

Para hacer Data Science la información debe pasar por unos procesos los cuales tienden a ser así:

#### -Preguntas problemáticas

Las cuales son preguntas que nos orientan a la resolución de un problema o una necesidad ejemplo de ello es ¿Cuál es la preferencia estilo, color y diseño en zapatillas deportivas de los clientes jóvenes entre 14 a 22 años?

#### -Recopilación de datos

La finalidad de este proceso es reunir datos de fuentes confiables que se relacionen a la finalidad con el fin de comenzar a tener tendencias, un ejemplo con el tema de las zapatillas es la recopilación de facturas de ventas de la empresa con respecto a los zapatos.

#### -Filtrado, procesamiento y la organización de los datos

Es el proceso que nos permite eliminar datos que dificulten el proceso de análisis y toma de decisiones, hay datos que pueden que pueden distorsionar la salida debido a no tener relación con el problema, además de tener datos ordenados nos



facilita llegar a una respuesta más veraz y certera con menos procesamiento, esto es importante en el tema de infraestructura y capacidad computacional. Ejemplo los nombres y apellidos de una factura no nos va permitir conocer las preferencias en el calzado de una persona, sin embargo si ordenamos por colores por tipos, por modelos, por cantidad de ventas y precios podemos llegar a mejores resultados.

#### -Modelos y algoritmos

Tienen la finalidad de que por medio de modelos estadísticos y algoritmos encontrar la solución al interrogante inicial, la interacción entre los datos permite realizar procesos de correlación, comparación y ejecución de altos flujos de datos que la mente humana no sería capaz de realizar, un modelo o algoritmo podría realizarlo en cuestión de segundos con una infraestructura computacional que lo respalde.

#### - Revisión de resultados y aprendizaje

Con la salida de los datos encontraremos respuesta en forma de estadísticas y gráficos que nos permitirán tener una visual de la salida de los datos, tener un panorama de las tendencias y posibles resultados que se acerquen a lo que buscamos.

#### -Toma de decisiones

Es el proceso en el cual ya teniendo nuestros resultados y habiendo revisado los datos de salida que nos arrojó el sistema tomaremos decisiones para conseguir el objetivo inicial del problema que queremos resolver.

### **Machine learning**

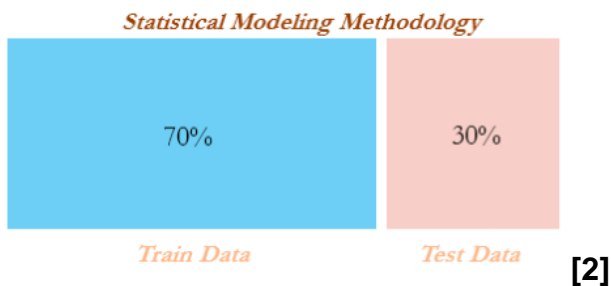
El Machine Learning es un campo de la informática que tiene la finalidad del aprendizaje automático de la maquina sin necesidad de estar programado para hacer una tarea sino a partir de modelos estadísticos embebidos o algoritmos en la maquina esta pueda comenzar a entender patrones y dar resultados a partir de simplemente datos de entrada. Esto lo hace la maquina realizando procesos de

prueba de error en varios ciclos hasta encontrar patrones que le permitan tener predicciones.

### Datos de entrenamientos y datos de prueba

Uno de las características principales del Machine Learning es que los datos se dividen en dos formas en datos de entrenamiento y datos de prueba. Los datos iniciales que se utilizan para optimizar y construir el modelo, entrenarlo se conocen como datos de entrenamiento, la calidad de nuestro modelo de Machine Learning depende de la calidad de los datos que le entreguemos en la entrada, por eso es importante poder realizar un buen filtrado y clasificación de la información si queremos resultados más veraces.

Los datos de prueba son datos que utilizamos para realizar un testeo de si nuestro modelo de entrenamiento funciona de acuerdo a las necesidades que estamos buscando, es importante porque nos define si el modelo es acertado o no. Es importante reservar una mayor cantidad de datos para la prueba que los de entrenamiento.



### Aprendizaje Supervisado

El aprendizaje supervisado es una rama del Machine Learning el cual se centra en la generación de patrones de aprendizaje por medio de la conexión entre las variables con los resultados conocidos y los dataset de datos etiquetados, esto se da funcionando alimentado los datos de muestra con varias características y el valor que debería tener el modelo en su salida. Lo importante de saber los datos que deberían salir es que ajusta los datos a procesar hasta que estos puedan tener cada vez una incertidumbre menor, un ejemplo de este tipo de aprendizaje es cuando se requiere predecir la relación entre las características que tiene un teléfono móvil por ejemplo su color, su diseño, los materiales, componentes electrónicos, software, cámara, sonido y prestaciones con la cantidad de teléfonos vendidos. Para que nos serviría para poder tener crear un teléfono que tenga una excelente aceptación de parte del consumidor.

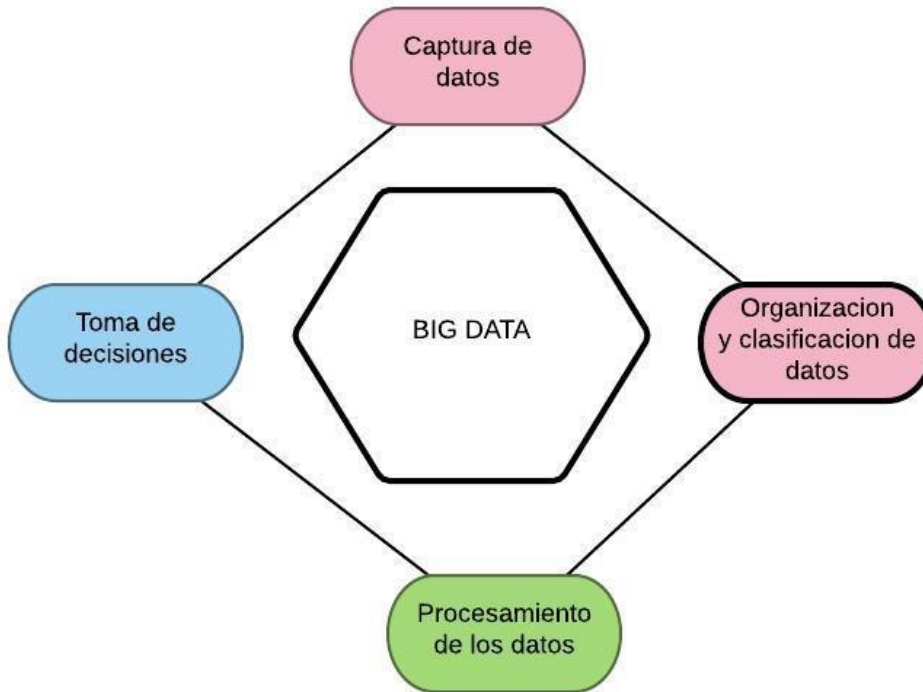
### Aprendizaje no supervisado

Para este tipo de Aprendizaje la maquina debe descubrir patrones ocultos y etiquetar los datos por medio de un algoritmo de aprendizaje no supervisado que es el que le indica cómo debe que buscar y hacer el etiquetado de la información, la importancia de este tipo de aprendizaje es que podemos encontrar patrones que no conocemos en grupos de datos no clasificados, esto es útil por ejemplo en los ataques cibernéticos a las grandes multinacionales o bases de datos gubernamentales, con un modelo de aprendizaje no supervisado podemos crear un algoritmos que nos tome datos de los intentos de ataques a nuestros sistemas, el comportamiento de estos ataques, las tendencias, las vulnerabilidades que busca, que proceso realiza para intentar vulnerar nuestra seguridad.

### Aprendizaje por refuerzo

Es otra rama del machine learning y se podría decir que es de las más complejas a comparación del aprendizaje supervisado y el no supervisado, debido a que es un tipo de aprendizaje que se da intentando mejorar continuamente el modelo aprovechando la interacciones que realice los datos cada vez que pasan nuevamente por el modelo, esto con el fin de hacerlo más robusto y mejor. Un ejemplo de esto es lo que se practica en algunos de automóviles autónomos, estos realizan por ejemplo cambios de dirección a determinadas velocidades, frenados fuertes y débiles, además de detección de elementos a su alrededor, con el fin de que cuando los autos ya estén en la calle identifiquen condiciones de las vías para llevar a los pasajeros en las vías, como obstáculos, desniveles, tráfico y parámetros climáticos, todo ellos buscando la gestión y optimización de giros en las calles, la seguridad de los pasajeros, el consumo de energía o combustibles y el uso de neumáticos. (Theobald, 2017)

## Big Data



[3]

En la actualidad estamos inundados de datos, desde las interacciones que tenemos en redes sociales, como en las compras que realizamos de diversos productos en cualquier cadena de supermercados, todos ellos constituyen billones de datos los cuales dan tendencia de nuestros hábitos y comportamientos diarios como consumidores, la labor del Big Data es tomar todo ese conjuntos de datos de gran volumen, capturarlos, clasificarlos y procesarlos mediante técnicas o herramientas tecnológicas. Esto es muy importante para la toma de decisiones, dado que podemos encontrar patrones en el comportamiento de las personas, del mercado etc. Esto se hace recopilando los datos y buscando en ellos tendencias los cuales ayuden a resolver preguntas. (Provost, 2013)

### Aplicaciones del Big Data

Publicidad

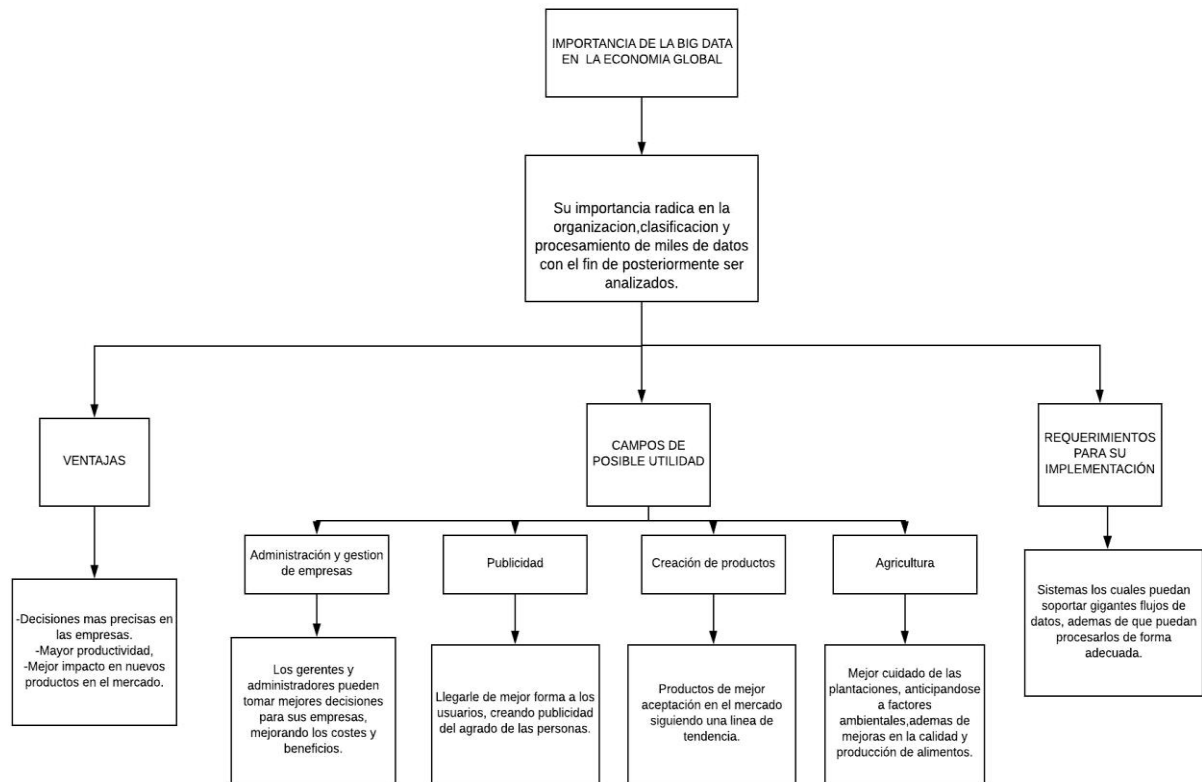
Por medio de los datos capturados de los consumidores por ejemplo compra de productos, redes sociales, costumbres etc. Se pueden crear oportunidades en campañas que impacten a los usuarios, llamando la atención y haciendo que estos piensen en la necesidad del producto que se requiere mostrar.

### Infraestructura tecnológica

Un ejemplo bueno para las infraestructura tecnológicas Big Data ayuda a conocer de mejor medida los puntos fuertes y puntos débiles de la infraestructura física y lógica de una organización, cambios que garanticen una optimización de la interacción de las estructuras y al mismo tiempo la reducción de costes, por ejemplo de las redes de internet con los dispositivos que se ajusten a las necesidades de una compañía, el ajuste de la tecnología que permita mayor productividad y menores costes en las empresas.

### Industria Retail

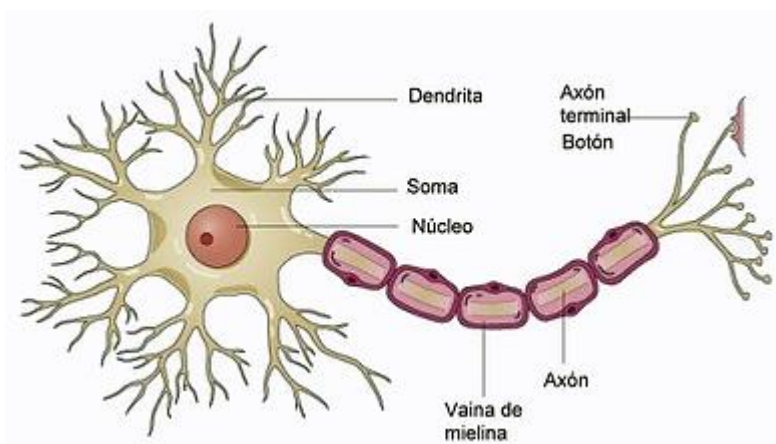
Es muy útil la aplicación del Big Data en la industria Retail, debido a que ayuda a que las compañías puedan conocer que productos tienen alta demanda en el negocio, que decisiones se pueden tomar para que crezca la ventas y mejorar los procesos internos para ajustarse a las necesidades del consumidor final, un ejemplo de ello es por ejemplo las bebidas hidratantes, con el Big Data la industria Retail puede saber cuáles bebidas son preferidas por el mayor grupo de clientes, que sabores son los favoritos, diseños de envases que llamen la atención del consumidor.



[4]

## Redes Neuronales artificiales –RNA

Las redes Neuronales artificiales son redes las cuales están creadas a partir de la simulación del funcionamiento de las neuronas humanas, como éstas interactúan en nuestro cerebro para ayudarnos a realizar tareas de nuestra vida cotidiana. Estas cómo se comporta frente a un impulso recibido en sus dendritas y como es su respuesta hacia las otras por medio de los axones.



[5]

La forma de trabajo de las neuronas artificiales es que con la suma de las entradas multiplicadas cada uno por sus respectivos pesos simula el impulso nervioso que tiene recibe cada neurona artificial, esto hace que internamente de la neurona artificial se procese los valores que recibe mediante unas funciones de activación que da como respuesta de salida que pasa a la siguiente neurona. Las redes neuronales se organizan en diferentes tipos de capas:

#### Capa de entrada

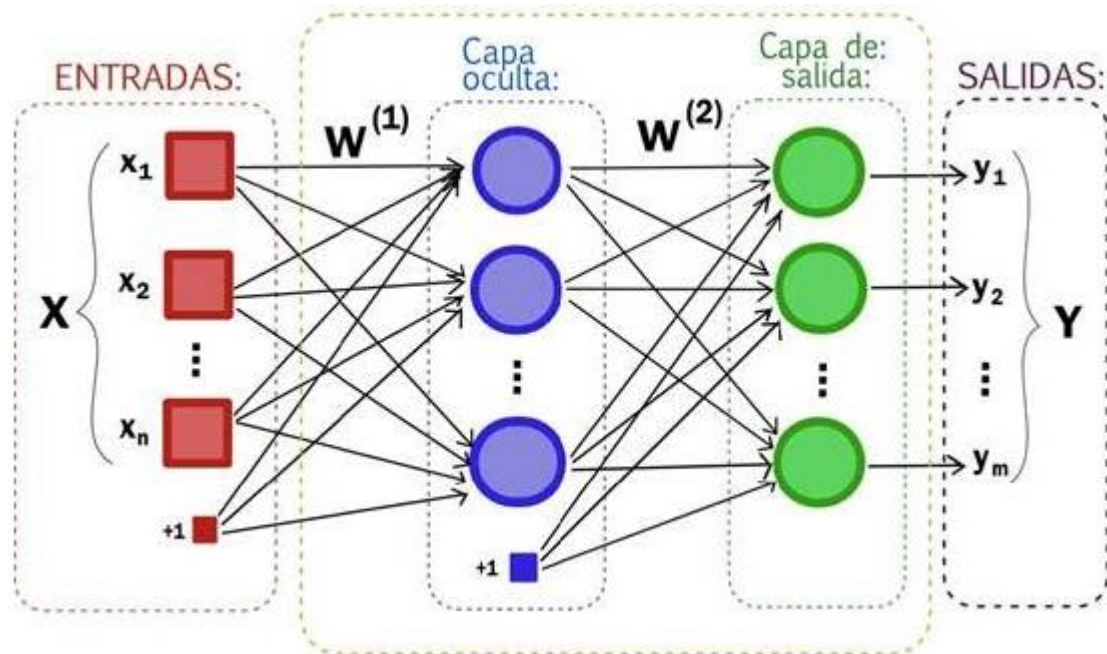
Es la capa en la cual se alimenta la red neuronal con los datos de entrada al sistema.

#### Capa Oculta

Es la capa donde se encuentra las interacciones no observables de la red neuronal y en esta se alojan las diferentes funciones de activación que procesan las unidades de entrada dando unidades respuesta para las siguientes capas.

#### Capa de salida

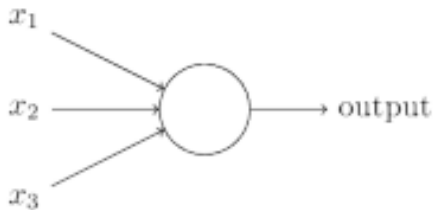
Son aquellas neuronas de nuestra red artificial que proporciona una respuesta del procesamiento que hubo a nuestra red neuronal.



[6]

## Perceptron

Los perceptrones son unidades neuronales en las redes neuronales las cuales permiten recibir múltiples entradas dando como respuesta una única salida binaria, el principal modelo que utiliza internamente es conocido función sigmoidea neuronal, este utiliza una función embebida que le permite seleccionar como lo hace una neurona de diferentes entradas la información y convertirla en una sola respuesta. (Nielsen, 2015)



[7]

## **APLICACIÓN DEL MACHINE LEARNING Y BIG DATA EN EL ANALISIS DE LA HUELLA DE ASFALTO EN CULTIVOS DE MAÍZ**

El proyecto en el cual trabaje para mi proyecto de grado fue a partir de un dataset de los datos capturados de las variables ambientales en un cultivo de maíz en el cual se media por medio de sensores de humedad, temperatura y revisión de los índices de severidad a de la infección por parte del patógeno llamado mancha de asfalto, la severidad fue determinada con ayuda de un espectrómetro portátil que entregaba firmas espectrales de plantas muestras dentro del cultivo, siempre las muestras fueron en las mismas plantas las cuales eran etiquetadas para conocer de cual planta se trataba.

### **DATASET**

A partir de los datos que se tomó de las plantas desde el día 08 de Agosto de 2019 hasta el día 17 de octubre del 2019, se construyó un dataset con la siguiente información de izquierda a derecha:

-La fecha: en que se tomó la muestra.

-Id de muestra: es decir como mencione cada planta de muestra siempre fue la misma y fue etiquetada.

-Descripción de los síntomas: Las características físicas que presentaba las hojas de las plantas.



-Observaciones microscópicas: Las cuales son observaciones encontradas en los microscopios de la Universidad Católica de Manizales en colaboración de estudiantes en proceso de proyecto de grado del Pregrado de Bacteriología.

-Severidad: Es el porcentaje de afectación que tiene la planta por el patógeno.

-Microorganismo: el cual fue se clasifico para que fuera el de estudio la *Phyllachora Maydis* conocida como Mancha de asfalto.

-Temperatura: La temperatura que marcaban los sensores en el día que se tomó las muestras.

-Temperatura promedio: La temperatura promedio calculado el día de la toma muestral.

-Temperatura Desviación: Variación o dispersión de la temperatura con la media.

-Humedad Relativa: La humedad relativa en el ambiente.

-Humedad Relativa Promedio: La humedad relativa promedio en el ambiente el día de la toma muestral.

-Humedad Relativa Desviación: La desviación estándar que tenía la humedad el día de la toma muestral.

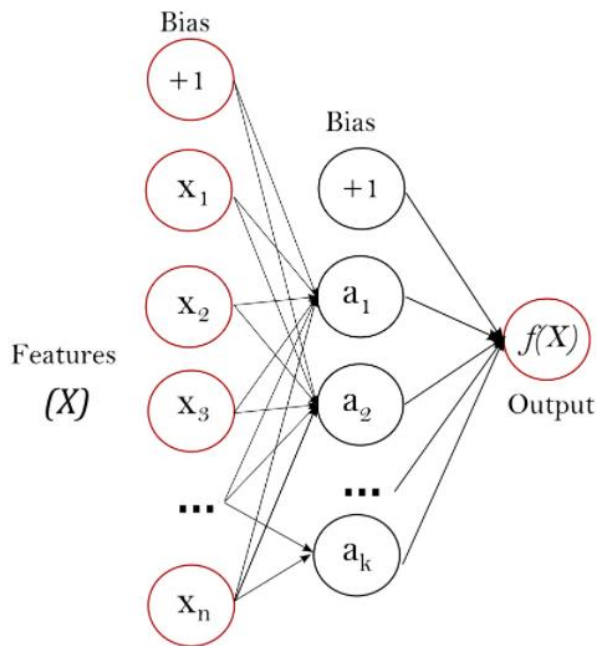
-Precipitación: La precipitación en el cultivo en milímetro por hora mm/h.

-Estado Fenológico: La cual es la etapa del periodo vegetativo en la que se encuentra el cultivo en el momento de la toma muestral.

FECHA	ID MUESTRA	DESCRIPCIÓN DE LOS SÍNTOMAS	OBSERVACIONES MICROSCÓPICAS	SEVERIDAD	MICROORGANISMO	TEMPERATURA	TEMPERATURA PROMEDIO	TEMPERATURA DESVIACIÓN	HUMEDAD RELATIVA	HUMEDAD RELATIVA PROMEDIO	HUMEDAD RELATIVA DESVIACIÓN	RECIPITACIÓN	ESTADO FENOLOGICO
08/08/2019	P1V81	Puntos negros brillantes y abultados	Ascas y ascosporas hialinas	15%	<i>Phyllochora maydis</i>	29	29,25	1,10	40	41	2,98	1	V6
08/08/2019	P2V81	Puntos negros brillantes y abultados	Ascas y ascosporas hialinas	15%	<i>Phyllochora maydis</i>	29	29,25	1,10	40	41	2,98	1	V6
08/08/2019	P3V81	Puntos negros brillantes y abultados	Ascas y ascosporas hialinas	15%	<i>Phyllochora maydis</i>	29	29,25	1,10	40	41	2,98	1	V6
08/08/2019	P4V82	Puntos negros brillantes y abultados	Ascas y ascosporas hialinas	15%	<i>Phyllochora maydis</i>	29	29,25	1,10	40	41	2,98	1	V6
08/08/2019	P5V82	Puntos negros brillantes y abultados	Ascas y ascosporas hialinas	15%	<i>Phyllochora maydis</i>	29	29,25	1,10	40	41	2,98	1	V6
08/08/2019	P6V82	Puntos negros brillantes y abultados	Ascas y ascosporas hialinas	15%	<i>Phyllochora maydis</i>	29	29,25	1,10	40	41	2,98	1	V6
08/08/2019	P7V83	Puntos negros brillantes y abultados	Ascas y ascosporas hialinas	15%	<i>Phyllochora maydis</i>	29	29,25	1,10	40	41	2,98	1	V6
08/08/2019	P8V83	Puntos negros brillantes y abultados	Estromas y ascas y ascosporas hialinas	15%	<i>Phyllochora maydis</i>	29	29,25	1,10	40	41	2,98	1	V6
08/08/2019	P9V83	Puntos negros brillantes y abultados	Estromas y ascas y ascosporas hialinas	15%	<i>Phyllochora maydis</i>	29	29,25	1,10	40	41	2,98	1	V6

## DESCRIPCION DEL ALGORITMO Y PROGRAMACION

## PERCEPTRON MULTICAPA (MLP)



[8]

El modelo que utilice en Python para realizar las predicciones de severidad de la planta bajo las condiciones climáticas y parámetros ambientales fue el modelo de Perceptron multicapa (MLP), el cual es un algoritmo de redes neuronales artificiales del tipo de aprendizaje supervisado el cual aprende una función:

$$f(.) : R^m \rightarrow R^0$$

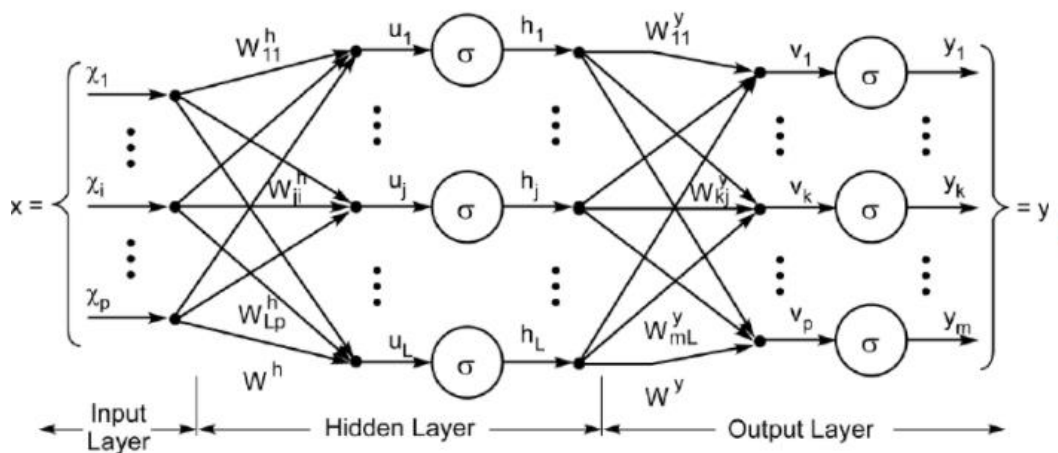
Donde esta función entrena un conjunto de datos los cuales tienen como denominación de que  $m$  en la ecuación es el número de dimensiones en la entrada y  $0$  es las dimensiones del sistema en su salida. En la capa de entrada tiene un conjunto de características  $X(n)$  y un objetivo o una meta puede aprender un aproximado lineal de regresión o clasificación esto con el fin de realizar ajustes al modelo para encontrar el objetivo buscado.

## REGRESOR DE PERCEPTRON MULTICAPA

Más específicamente el modelo de Perceptron Multicapa que se utiliza es el regresor de Perceptron Multicapa, el cual es un modelo estadístico y matemático que se basa en la técnica de descenso de gradientes estocásticos. El cual es un

modelo ampliamente utilizado en proyectos de inteligencia artificial y de Machine Learning el cual está conformado por:

- UNA CAPA DE ENTRADA: capa en la cual nuestro Perceptron o Neurona artificial recibe la señal de entrada y en donde vamos a imprimir las características que necesitamos de nuestro sistemas, teniendo en cuenta el objetivo del mismo, además de que allí ajustamos un BIAS "Sesgo" el cual es un parámetro en nuestros nodos para evitar que el algoritmo no aprenda de forma incorrecta al carecer de información útil o necesaria para el procesamiento en el sistema esto así evitando desajustes, otro punto importante en la capa de entrada es un parámetro llamado el peso neuronal el cual es un parámetro útil al tener varias entradas debido a que con ellas sabremos cual característica es más importante que las demás, esto es importante porque dependiendo a cual característica le demos más peso así se verá reflejada la respuesta de la red neural .
- UNAS CAPAS OCULTAS: Las cuales serán modificadas a la necesidad el usuario, es un punto muy importante a tener en cuenta que si nuestro objetivo es encontrar respuesta no lineales debemos de tener más de una capa oculta, esto aumentara el nivel de complejidad y de computación pero ajustara mejor nuestro sistema a sistemas no lineales, mientras un sistema de un solo Perceptron es útil para respuestas lineales,



[9]

- CAPA DE SALIDA: la capa en la cual tendremos nuestra respuesta a la interacción entre perceptrones y parámetros de la red neuronal es importante tener en cuenta que las redes interactúan tanto de las capas de entrada a la oculta y posteriormente a la de salida como en sentido contrario, este en

su estructura final tendrá embebida el modelo de gradiente descendente estocástico, el cual nos permitirá hacer los ajustes con los datos de entrenamiento que le entreguemos y devolverá un respuesta ajustando nuevamente los pesos y las Bias con el fin de que en la siguiente interacción de datos este nos entregue una respuesta más cercana a la salida que busquemos, en varias interacciones hasta que encuentre una respuesta mínima, es decir se acerque lo mas poco a la respuesta. Este es el modelo de **descenso de gradiente estocástico**:

$$\frac{\partial F}{\partial w_0} = -2 \sum_{i=1}^n (y_i - f(W; x_i))$$

Este algoritmo se usa principalmente por su bajo costo computacional, es más sencillo de su uso a comparación que otros modelos, es muy útil para encontrar tendencias y tener predicciones sobre situaciones complejas en tiempo real.

### Por qué Python?

Utilice para la programación de mi proyecto el IDE Spyder 3.3.3 con motor Python debido a que es un lenguaje orientado a ambientes de investigación científica, realizar pruebas dinámicas e interactivas, además de que es muy flexible para cualquier desarrollo orientado al Machine Learning, además de que es muy amigable con nosotros los principiantes al momento de programar, debido a la gran cantidad de librerías con funciones que se pueden llamar para ser utilizadas en nuestros proyectos sobre todo desde una orientación matemática y estadísticas, en comparación de otros lenguajes como C o C++ es menos complejo y más flexible, se acomoda más fácil a las necesidades del programados, sin embargo a tener en cuenta que los últimos mencionados son mucho más robustos dado su complejidad. Existen lenguajes mucho más fáciles para poder realizar la temática con que fue enfocado este proyecto de grado sobre todo para el cálculo de correlaciones y de las redes neuronales como es R, es una herramienta muy amada por los científicos pues aumenta la productividad y es menos complejo sin embargo es un lenguaje que se basa más en los resultados que en los procesos, entonces cuando se deba migrar a otros lenguajes o concatenar entre dos o más tendremos dificultades graves para entender y hacer ajustes necesarios.

## CODIGO DE PROGRAMACION PYTHON

```
"""
```

```
Created on Sat Mar 20 19:20:46 2021
```

```
Jhon Sabogal
```

```
"""
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy.stats import pearsonr
from sklearn.neural_network import MLPRegressor
```

```
File= pd.ExcelFile('Mancha.xlsx')
print(File.sheet_names)
```

```
df=File.parse('datos')
```

```
df2=File.parse('datos')
```

```
DATA= df[["DÍA", "SEVERIDAD", "TEMPERATURA", "HUMEDAD RELATIVA"]]
```

```
print(DATA)
```

```
DIA= df[["DÍA"]]
SEVERIDAD= df[["SEVERIDAD"]]
TEMPERATURA= df[["TEMPERATURA"]]
HUMEDAD= df[["HUMEDAD RELATIVA"]]
```

```
corr, _ = pearsonr(SEVERIDAD, TEMPERATURA)
print('Correlacion de SEV vs TEMP: %.3f' % corr)
```

```
corr, _ = pearsonr(SEVERIDAD, HUMEDAD)
print('Correlacion de SEV vs HUM: %.3f' % corr)
```

```
x=df2['DÍA']
y=df2['SEVERIDAD']
u=df2['TEMPERATURA']
v=df2['HUMEDAD RELATIVA']
```

```

plt.figure()
plt.subplot(2,2,1)
plt.plot(x,y)
plt.scatter(x, y, marker='o' ,color ='red')
plt.xlabel('Dias')
plt.ylabel('Phyllachorae Maydis')
plt.title('TIEMPO VS ENFERMEDAD')
plt.grid(True)

```

```

plt.subplot(2,2,2)
plt.plot(x,u)
plt.scatter(x, u, marker='o' ,color ='green')
plt.xlabel('Dias')
plt.ylabel('Temperatura Cº')
plt.title('TIEMPO VS TEMPERATURA')
plt.grid(True)

```

```

plt.subplot(2,2,3)
plt.plot(x,v)
plt.scatter(x, v, marker='o' ,color ='yellow')
plt.xlabel('Dias')
plt.ylabel('Humedad')
plt.title('TIEMPO VS HUMEDAD')
plt.grid(True)
plt.figure()

```

```
X=x[:,np.newaxis]
```

```

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y)

```

```

mlr=MLPRegressor(solver='lbfgs',alpha=1e-
5,hidden_layer_sizes=(3,3),random_state=1)
mlr.fit(X_train,y_train)
print(mlr.score(X_train,y_train))

```

```

prediccion=mlr.predict([[80]])
print("Prediccion en dias:", prediccion)

```

## DESCRIPCION DE HALLAZGOS Y CONCLUSIONES

El proyecto se originó en el municipio de Palestina, Caldas en la sede campestre de la Universidad De Caldas donde nos permitieron amablemente realizar el estudio correspondiente, allí se encontraba unos cultivos de Maíz en el cual nos sirvió para hacer el seguimiento del crecimiento de las plantas, al mismo tiempo como en la vida de la plantación iba creciendo diferentes patógenos, más específicamente en este caso se reviso fue el de la Mancha de Asfalto (PHYLLACHORAE MAYDIS), además de como a través del paso de los días estos interactuaban entre si además como los factores ambientales de ese rango de días hacia que cambiara el comportamiento de ambas especies.

Con ello recolectamos un Dataset de datos, se realizó la carga de datos a partir del Excel y en el algoritmo de programación realizamos la clasificación de los datos que iban a ser útiles para el procesamiento y encontrar la respuesta que buscábamos.

```
13 File= pd.ExcelFile('Mancha.xlsx')
14 print(File.sheet_names)
15
16 df=File.parse('datos')
17
18 df2=File.parse('datos')
19
20 DATA= df[["DÍA", "SEVERIDAD", "TEMPERATURA", "HUMEDAD RELATIVA"]]
21
22 print(DATA)
```

Esta fue la respuesta de dicha clasificación del Dataframe en un vector se nombró Data, esta fue su respuesta.

	DÍA	SEVERIDAD	TEMPERATURA	HUMEDAD RELATIVA
0	1	0.15	29.0	40
1	1	0.15	29.0	40
2	1	0.15	29.0	40
3	1	0.15	29.0	40
4	1	0.15	29.0	40
5	1	0.15	29.0	40
6	1	0.15	29.0	40
7	1	0.15	29.0	40
8	1	0.15	29.0	40
9	2	0.15	31.5	35
10	2	0.15	31.5	35
11	2	0.15	31.5	35
12	2	0.15	31.5	35
13	2	0.15	31.5	35
14	2	0.15	31.5	35
15	2	0.15	31.5	35
16	2	0.15	31.5	35
17	2	0.15	31.5	35
18	6	0.15	31.0	39
19	6	0.15	31.0	39
20	6	0.03	31.0	39
21	6	0.03	31.0	39
22	6	0.15	31.0	39
23	6	0.03	31.0	39
24	6	0.03	31.0	39
25	7	0.03	29.5	45
26	7	0.03	29.5	45
27	7	0.03	29.5	45
28	7	0.03	29.5	45
29	9	0.03	28.2	42

Luego de ello se revisó estos datos para entender la relación que hay entre el aumento de la severidad de la enfermedad, temperatura y humedad influenciadas por los cambios de climas en la región, para encontrar una relación se realizó un algoritmo matemático y estadístico llamada Coeficiente de Correlación de Pearson:

$$r_{xy} = \frac{\sum z_x z_y}{N}$$



Esta correlación se calcula realizando la sumatoria de la desviación estándar de una variable x y la desviación estándar de una variable y sobre el número de datos que tendremos en nuestro sistema. Esto se hizo de la siguiente manera llamando una librería que nos traía el modelo de Scipy:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy.stats import pearsonr
```

Luego de ello se llamó la función `Corr, _` la cual nos permitió el respectivo cálculo de la correlación Vs dos variables:

- SEVERIDAD VS TEMPERATURA
- SEVERIDAD VS HUMEDAD

```
24 DIA= df[["DÍA"]]
25 SEVERIDAD= df[["SEVERIDAD"]]
26 TEMPERATURA= df[["TEMPERATURA"]]
27 HUMEDAD= df[["HUMEDAD RELATIVA"]]
28
29
30
31 corr, _ = pearsonr(SEVERIDAD, TEMPERATURA)
32 print('Correlacion de SEV vs TEMP: %.3f' % corr)
33
34 corr, _ = pearsonr(SEVERIDAD, HUMEDAD)
35 print('Correlacion de SEV vs HUM: %.3f' % corr)
~
```

Entregándonos estas respuestas:

```
[152 rows x 4 columns]
Correlacion de SEV vs TEMP: -0.337
Correlacion de SEV vs HUM: 0.334
```

Para entender de mejor forma este resultado nos toca remontarnos a que dice Pearson de la respuesta a los coeficientes entre dos variables:

- Los coeficientes siempre estarán entre -1 y 1
- Cuando un coeficiente se acerca a 0 esto nos indicara que los dos eventos no tienen relación alguna.
- Cuando un coeficiente 1 significa que las variables tienen una relación fuerte la una de la otra.
- Cuando un coeficiente -1 significa que las variables tienen una relación inversa fuerte la una de la otra.
- Cuando un coeficiente esta entre 0,1 a 1 este tendrá una relación débil entre las variables.
- Cuando un coeficiente esta entre -0.1 a -1 tendrá una relación inversa débil entre las variables. (Boslaught, 2012)

### **CONCLUSION DE LA CORRELACION:**

Entonces entendiendo las premisas que podemos entender:

Entre menor sea la temperatura mayor puede ser de que se aumente el grado de afectación del patógeno a la planta es decir es inversamente proporcional.

Entre mayor sea la Humedad mayor puede ser que aumente el porcentaje de afectación del patógeno a la planta.

Esto lo podemos ver más fácilmente en la respuesta grafica que diseñe en el Script para ver visualmente estos comportamientos:

```

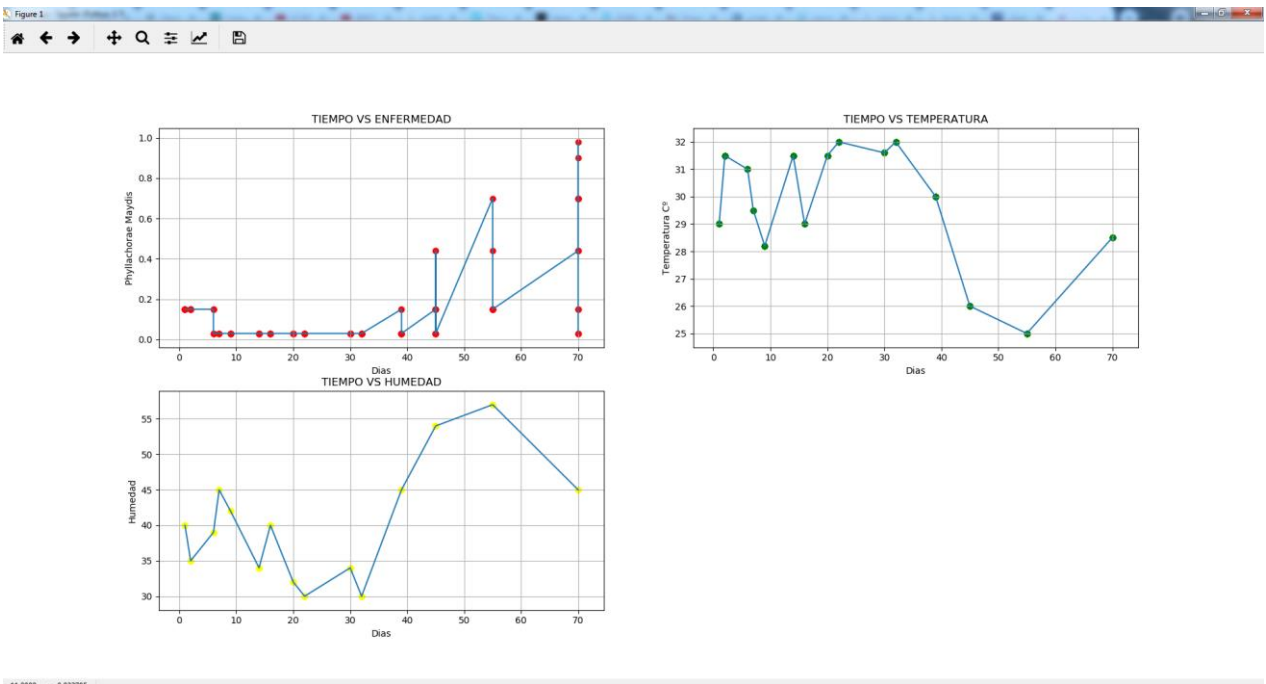
plt.figure()
plt.subplot(2,2,1)
plt.plot(x,y)
plt.scatter(x, y, marker='o' ,color = 'red')
plt.xlabel('Dias')
plt.ylabel('Phyllachorae Maydis')
plt.title('TIEMPO VS ENFERMEDAD')
plt.grid(True)

plt.subplot(2,2,2)
plt.plot(x,u)
plt.scatter(x, u, marker='o' ,color = 'green')
plt.xlabel('Dias')
plt.ylabel('Temperatura Cº')
plt.title('TIEMPO VS TEMPERATURA')
plt.grid(True)

plt.subplot(2,2,3)
plt.plot(x,v)
plt.scatter(x, v, marker='o' ,color = 'yellow')
plt.xlabel('Dias')
plt.ylabel('Humedad')
plt.title('TIEMPO VS HUMEDAD')
plt.grid(True)
plt.figure()

```

Entregándonos las siguientes graficas:



Si tenemos analizamos las gracias podremos encontrar que la severidad fue de comportamiento estable hasta el día 30 donde comenzó a tener picos más altos, entonces que nos puede mostrar la graficas de la temperatura en esos días, la temperatura disminuyo desde el día 30 desde los 32 grados hasta temperaturas mínimas de 25 grados esta disminución según nuestro de modelo de Pearson favorece el crecimiento del patógeno. Ahora lo mismo sucede con la humedad desde el día 30 la humedad aumento y la severidad aumento, entonces con esto gráficamente podemos demostrar que nuestro modelo cumple las expectativas y objetivos trazados en el proyecto.

## Red Neuronal

En este espacio del Script se tomó los datos que teníamos y se realizó una clasificación entre datos de entrenamiento y datos de prueba:

Se determinó la estructura de las capas de la red neuronal y se implementó el modelo de Perceptron Multicapas, el cual en su capa de salida nos daba una interacción nueva entre la nueva entrada de datos ajustando los pesos de cada uno de los nodos para que cada predicción fuera más exacto.

```
71 X=x[:,np.newaxis]
72
73
74 from sklearn.model_selection import train_test_split
75 X_train, X_test, y_train, y_test = train_test_split(X, y)
76
77 mlr=MLPRegressor(solver='lbfgs',alpha=1e-5,hidden_layer_sizes=(3,3),random_state=1)
78 mlr.fit(X_train,y_train)
79 print(mlr.score(X_train,y_train))
80
81 prediccion=mlr.predict([[90]])
82 print("Prediccion en días:", prediccion)
```

Las predicciones se hacían cambiando en la variable predicción en la imagen se puso el día 90, pero se puede ajustar a días futuros o días formulados en el dataset, un ejemplo es yo lo tome para el día 90 que es 20 días después de los datos que tenemos el dataset y el predice cómo será la severidad en la plantación posiblemente en esos días

```
0.41530868651899133
Prediccion en días: [0.47612631]
```

Igualmente podemos revisar la fiabilidad de nuestra red revisando como fue en días que sí estuvo en el dataset por ejemplo utilizara como fue en el día 2:

```

[ datos ]
  DÍA SEVERIDAD TEMPERATURA HUMEDAD RELATIVA
0     1     0.15      29.0         40
1     1     0.15      29.0         40
2     1     0.15      29.0         40
3     1     0.15      29.0         40
4     1     0.15      29.0         40
5     1     0.15      29.0         40
6     1     0.15      29.0         40
7     1     0.15      29.0         40
8     1     0.15      29.0         40
9     2     0.15      31.5         35
10    2     0.15      31.5         35
11    2     0.15      31.5         35
12    2     0.15      31.5         35
13    2     0.15      31.5         35
14    2     0.15      31.5         35
15    2     0.15      31.5         35
16    2     0.15      31.5         35
17    2     0.15      31.5         35

```

El cual cómo podemos ver la dataset fue del 15% de afectación a la planta y esta fue la respuesta de la red neuronal:

**0.4154462240804542**

**Prediccion en días: [0.15043195]**

Igualmente del 15% en nuestro predictor.

## CONCLUSIONES DE LA RED NEURONAL

La red neuronal entrego respuestas satisfactorias a las predicciones que se le propusieron a partir de los datos entregados, el espíritu del BIG DATA es tener Dataframes de miles de datos para perfeccionar nuestra red neuronal, si se pudiera tener más datos sería mucho más exacto y estable, sin embargo esto conllevaría mejorar la infraestructura tecnológica para soportar el sistema.

Al tener la oportunidad de conocer un poco de las redes neuronales artificiales se comienza a dimensionar lo impresionante que es el cerebro humano, tenemos en nuestra cabeza una unidad casi infinita de computo, con energía limpia y con capacidades infinitas. Para el procesamiento de datos que realiza un cerebro cada segundo se necesitaría una unidad computacional tan poderosa que sería casi

imposible de construir, el Machine Learning, la IA y el BigData revolucionan cada día de nuestra vida y son técnicas que harán que evolucionemos teniendo mejor calidad de vida, hará mas sostenible la vida a nuestro planeta y nos dará un futuro del que no se había pensado.

## **TABLA DE REFERENCIA DE IMÁGENES.**

**[1] FOTO DE PLANTA DE MAIZ EN PALESTINA DE MI PROPIA AUTORIA**

**[2] IMAGEN DATOS ENTRENAMIENTO Y PRUEBA Statistics for Machine Learning by Pratap Dangeti**

**[3] DIAGRAMA DE FLUJO BIG DATA DE MI PROPIA AUTORIA**

**[4] MAPA CONCEPTUAL BIG DATA DE MI PROPIA AUTORIA**

**[5] NEURONA HUMANA <https://bebrainid.wixsite.com/bebrain/neuronas>**

**[6] REDES NEURONALES ARTIFICIALES**  
**<https://sites.google.com/site/mayinteligenciartificial/unidad-4-redes-neuronales>**

**[7] PERCEPTRON**  
**<https://static.latexstudio.net/article/2018/0912/neuralnetworksanddeeplearning.pdf>**

**[8] PERCEPTRON MULTICAPA [https://scikit-learn.org/stable/modules/neural\\_networks\\_supervised.html](https://scikit-learn.org/stable/modules/neural_networks_supervised.html)**

**[9] RED NEURAL ESTRUCTURA. [https://www.researchgate.net/figure/The-structure-of-a-multilayer-perceptron-neural-network\\_fig3\\_241347660](https://www.researchgate.net/figure/The-structure-of-a-multilayer-perceptron-neural-network_fig3_241347660)**

## **Referencias**

Boslaught, S. (2012). *Statistics in a Nutshell*. O'Reilly.

Climate & CAC. (2015). *ccacoalition.org*. Obtenido de <https://www.ccacoalition.org/en/activity/open-agricultural-burning>

Forbes. (2021). ¿Cómo se convertirá Colombia en el Silicon Valley de América Latina? *Emprendedores*.

- Frattini, S. (2019). *https://agronegocios.uniandes.edu.co/*. Obtenido de <https://agronegocios.uniandes.edu.co/2019/08/31/agricultura-de-precision-el-futuro-del-agro-colombiano/>
- Infoagro. (s.f.). *Infoagro*. Obtenido de <https://www.infoagro.com/herbaceos/cereales/maiz.htm#:~:text=El%20ma%C3%ADz%20llega%20a%20soportar,temperaturas%20de%2020%20a%2032%C2%BAC.>
- Lages, G. (2018). *Hotmart Blog*. Obtenido de <https://blog.hotmart.com/es/que-es-data-science/#>
- Nielsen, M. (2015). *Neural Networks and Deep Learning*.
- Pereyda–Hernández, J. (2008). *Etiology and management of tar spot (Phyllachora maydis Maubl.) of maize in Guerrero State, México*. Iguala, Guerrero. Carretera Iguala–Tuxpan, km 2.5.: CEIGUA/INIFAP.
- Provost, F. (2013). *DATA SCIENCE FOR BUSINESS*. O'REILLY.
- R, H. (2015). Escala Diagramática de Severidad para el Complejo Mancha de Asfalto del Maíz. *Revista mexicana de fitopatología*. Vol. 33, 1.
- Rodriguez, D. (2020). *Lifeder.com*. Obtenido de <https://www.lifeder.com/factores-ambientales/>
- Semana. (2020). Exportaciones agrícolas de Colombia llegarían a US\$7.400 millones en 2020. *Dinero*.
- Theobald, O. (2017). *MACHINE LEARNING FOR ABSOLUTE BEGINNERS*.