

**UNIVERSIDAD CATÓLICA DE HONDURAS CAMPUS  
JESÚS SACRAMENTADO**

**Integrantes:**

**Ramón Humberto Ventura Rodríguez**

**Heidy Nohemy Lemus Ramos**

**Will Alejandro Lara Mejía**

**Manuel Enrrique Aguilar Reyes**

**Abel Enrique Consuegra Zuniga**

**Catedrático:**

**Ing. Héctor Sabillon**

**Tema:**

**Árboles de decisión y bosques aleatorios**

**Clase:**

**Inteligencia artificial**

**I Periodo 2020**

**Siguatepeque Comayagua**

## Contenido

Contenido de ilustraciones .....	2
Introducción.....	3
Objetivos .....	4
Manual de desarrollo de proyecto .....	5
Herramientas utilizadas para el desarrollo.....	5
Jupyter Notebook .....	5
Roles de los integrantes .....	5
Responsabilidades .....	6
Reunión .....	6
Plan de trabajo.....	7
Trello .....	7
Sistema de control de versiones.....	8
GitHub.....	8
Link del Repositorio En GitHub.....	9
Ejemplo visual de las libretas.....	10
Árbol de decisión.....	10
Bosques aleatorios.....	11
Conclusión.....	12
Información consultada .....	13
Puntaje .....	13

## Contenido de ilustraciones

<i>Ilustración 1 Jupyter Notebook</i>	5
<i>Ilustración 2 Tabla de reunión</i>	7
<i>Ilustración 3 Plataforma en Trello</i>	8
<i>Ilustración 4 Rama en GitHub</i>	9
<i>Ilustración 5 Commit en GitHub</i>	9
<i>Ilustración 6 Teoría de los árboles de decisiones</i>	10
<i>Ilustración 7 Código de los árboles de decisiones</i>	10
<i>Ilustración 8 teoría de los Bosques aleatorios</i>	11
<i>Ilustración 9 Código de los Bosques aleatorios</i>	11

## **Introducción**

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.

A continuación se da una breve explicación de lo que son los arboles de decisiones y bosques aleatorios los cuales son algoritmos de machine learning, los cuales consisten en decisiones con respecto a los arboles este divide un nodo de decisión principal y los cuestiona, así sucesivamente va seleccionando el mejor nodo. Respecto a los bosques es un algoritmo de aprendizaje supervisado que, crea un bosque y lo hace de alguna manera aleatorio. Para decirlo en palabras simples: el bosque aleatorio crea múltiples arboles de decisión y los combina para obtener una predicción más precisa y estable.

## **Objetivos**

1. Ampliar los conocimientos sobre Data Science y los diferentes métodos de compresión de datos.
2. Fortalecer las habilidades de investigación, análisis de la información, capacidad de síntesis, análisis y programación, desarrollo de informe ejecutivo y transmisión de conocimiento en la temática asignada.

## Manual de desarrollo de proyecto

### Herramientas utilizadas para el desarrollo

#### *Jupyter Notebook*

El Jupyter Notebook es una aplicación web de código abierto, desarrollada utilizando lenguaje HTML agnóstico que permite crear, compartir y editar documentos en los que se puede ejecutar código python, hacer anotaciones, insertar ecuaciones, visualizar resultados y documentar funcionalidades.



*Ilustración 1 Jupyter Notebook*

Esta aplicación está diseñada generalmente para tener una compatibilidad avanzada con Python, Markdown e incluye la posibilidad de exportar documentos hechos con la herramienta a otros formatos.

Generalmente esta herramienta es utilizada para el aprendizaje del lenguaje de programación Python, la limpieza y transformación de datos científicos, la simulación numérica, el modelado estadístico entre muchas otras áreas.

### Roles de los integrantes

Sabemos que un rol es un papel que un individuo o cosa determinada juega en un determinado contexto, como en este caso en el proyecto de este sitio web.

- **Coordinador:** Como su palabra lo dice coordinar el trabajo del equipo tal como asignar tareas, definir días de reunión y de la misma manera estar supervisando y revisando lo que cada integrante tiene delegado para trabajar en dicho proyecto. Así mismo el coordinador define en cuanto tiempo el integrante debe de entregar los avances y supervisar que este se creó de la manera correcta.

- **Integrantes:** Cada integrante tiene delegado investigar acerca de los temas del proyecto y tratar de entenderlo, luego de esto podría comenzar a desarrollar las libretas en Jupyter Notebook.

### **Responsabilidades**

- **Nohemy Lemus (coordinador):**
  1. Coordinar las reuniones y tareas designadas a cada uno de los miembros tomando la palabra en cada reunión
  2. Se les asigna a los integrantes leer acerca de los temas asignados.
  3. Define las tareas a cada integrante del grupo.
  4. Crea los archivos necesarios para el desarrollo de las libretas.
- **Will Lara(integrante):**
  1. Ayuda al desarrollo de la libreta con respecto a los bosques aleatorios.
  2. Trabaja con documentación de los bosques aleatorios
- **Manuel (integrante):**
  1. Ayuda principalmente en el desarrollo de los bosques aleatorios.
- **Ramon Ventura (integrante):**
  1. Ayuda en la documentación de los árboles de decisión, principalmente en los árboles de clasificación
  2. Crea lo que es el árbol de clasificación
- **Abel Consuegra**
  1. Ayuda con la documentación de los árboles de decisión, principalmente en los de regresión
  2. Crea lo que es el árbol de regresión

### **Reunión**

Se realizó una reunión la fecha 10 de abril del 2020 con el objetivo de aclarar algunos puntos a cada integrante del equipo. Solamente se realizó una reunión dado el hecho que en la plataforma de Trello se estaba comunicando cada desarrollo que los integrantes del grupo iban haciendo.

También cabe mencionar que cada integrante se comunicaba de forma independiente con el coordinador al tener una duda, esta comunicación surgía por medio de la plataforma de WhatsApp

Cada integrante del equipo avanzaba individualmente y solo comunicaba cuando este terminaba su parte, o cuando estaba apoyando en alguna otra actividad.

Fecha	Propuestas y acuerdos
10-04-2020 02 horas	<ol style="list-style-type: none"><li>1. Presentación de los avances de cada integrante.</li><li>2. Cada integrante explico la parte de desarrollo que había realizado</li><li>3. Se explico la documentación encontrada y la que se va a presentar en el proyecto.</li><li>4. A cada integrante se le asigno lo que va a tener que desarrollar al presentar el proyecto.</li><li>5. El o los integrantes que conocían más acerca del tema trataron de explicar cómo funcionaban tanto los árboles de decisión como los bosques aleatorios.</li></ol>

Ilustración 2 Tabla de reunión

## Plan de trabajo

Decidimos utilizar para el plan de trabajo la herramienta:

### Trello

Es una herramienta que sirve para organizar tareas/proyectos en un tablero (boards) al que le podemos asignar el nombre del proyecto que estamos trabajando.

Este funciona de manera que cada integrante puede hacer una conversación o hacer observaciones sobre el proyecto. También se definen que partes del proyecto están en proceso, cuales son las tareas y que partes son las que ya están terminadas.

Como ejemplo tomamos en cuenta estas etiquetas o tablas para ir coordinando el trabajo de nuestro proyecto:



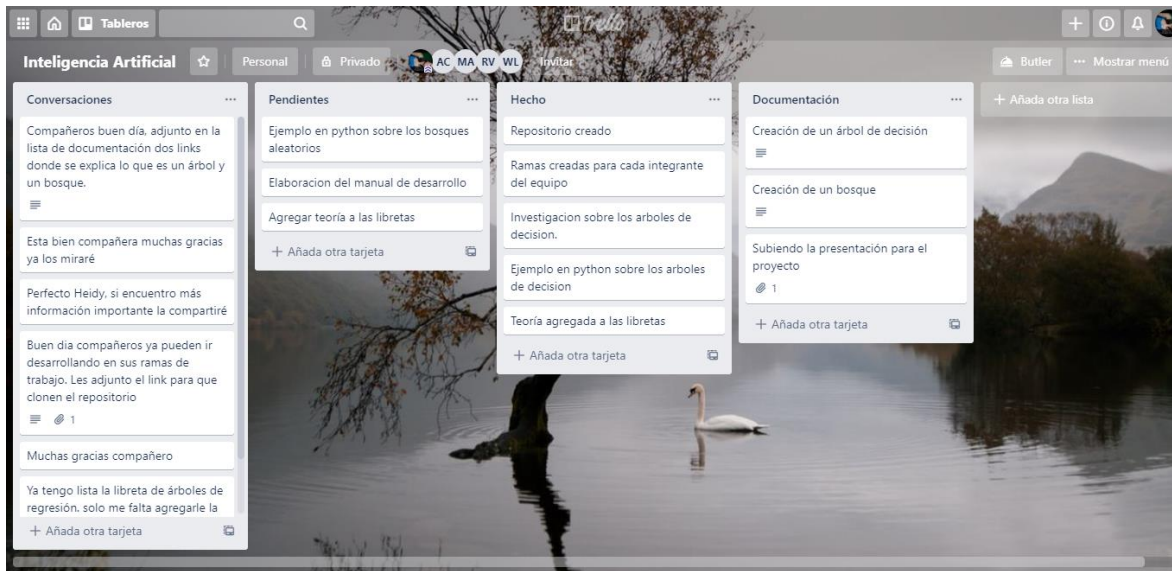


Ilustración 3 Plataforma en Trello

Link del tablero de Trello: [Enlace para agregarse al tablero](#)

## Sistema de control de versiones

### GitHub

Es una plataforma de desarrollo colaborativo de software para alojar proyectos utilizando el sistema de control de versiones Git. El código se almacena de forma pública, aunque también se puede hacer de forma privada, creando una cuenta de pago.

Esta herramienta funciona de forma que cada persona tenga una rama, y esta tiene que ir actualizando sus actualizaciones. Nos ayuda a tener los documentos de forma más ordenada y que de igual manera se viera cual era el avance de cada integrante.

En el proyecto de desarrollo de portales se creó una rama para cada integrante y una adicional para así ir compartiendo todos los avances y los demás poder actualizar las ramas con estas actualizaciones como vemos a continuación.

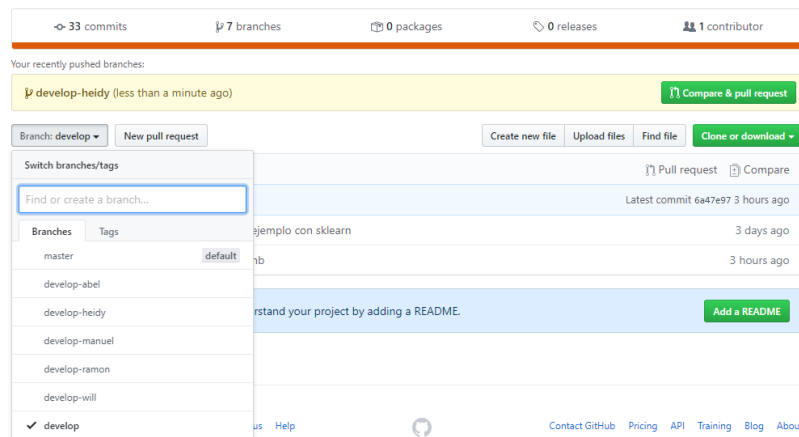


Ilustración 4 Rama en GitHub

Así mismo se ven las actualizaciones o commit que los integrantes han hecho:

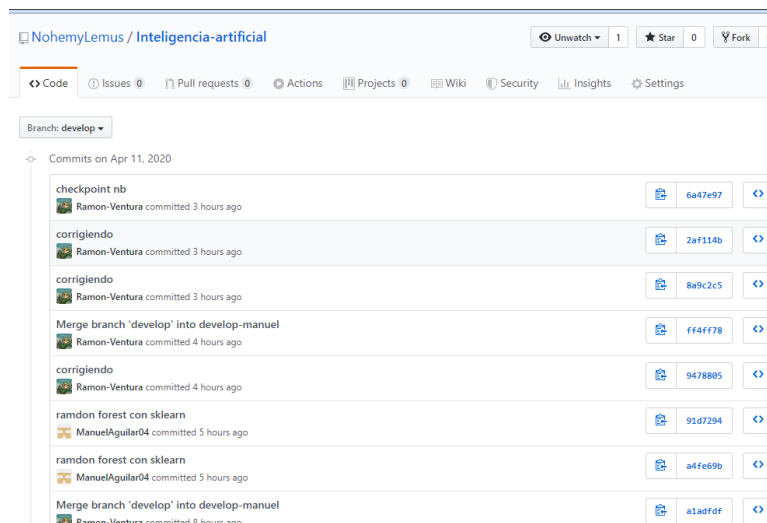


Ilustración 5 Commit en GitHub

Link del Repositorio En GitHub: [Repositorio del proyecto acerca de árboles de decisión y los bosques aleatorios](#)

## Ejemplo visual de las libretas

### Árbol de decisión

El árbol de decisión se realizó de tal manera que primeramente se explica lo que son, como se clasifica y como funciona, después se ve cómo se cargan los datos, la limpieza de lo mismo, y por último se empieza a hacer los que es el árbol.

Mencionar también que las funciones aplicadas en los arboles de decisiones también están especificadas o explicadas.

### Teoria

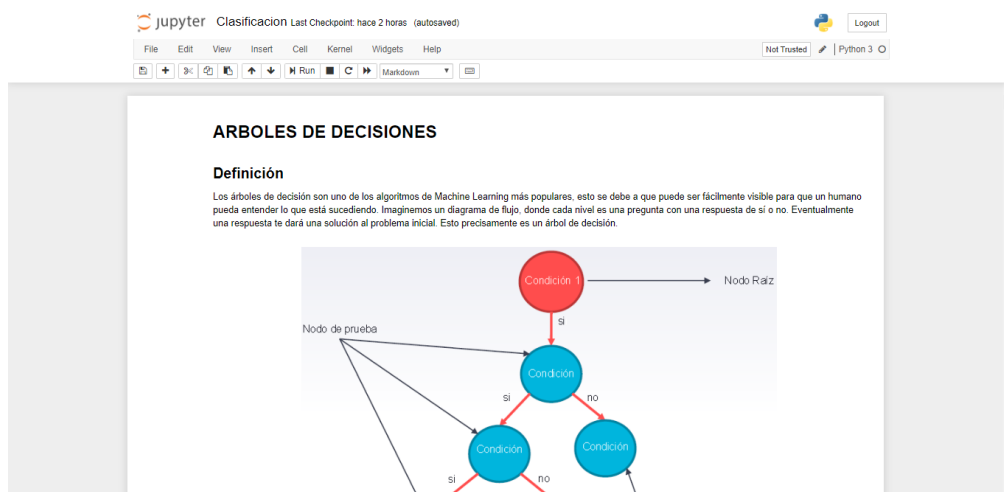


Ilustración 6 Teoria de los arboles de decisiones

### Código

```
jupyter Clasificación Last Checkpoint: hace 2 horas (autosaved)
File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3
+ - * / % < > Run C Markdown

Cargamos el dataframe

In [23]: df = pd.read_csv('../data/Iris.csv')
In [24]: df.head()
Out[24]:
  id  sepal_length  sepal_width  petal_length  petal_width  species
0  1         5.1         3.5         1.4         0.2  iris-setosa
1  2         4.9         3.0         1.4         0.2  iris-setosa
2  3         4.7         3.2         1.3         0.2  iris-setosa
3  4         4.6         3.1         1.5         0.2  iris-setosa
4  5         5.0         3.6         1.4         0.2  iris-setosa

Se procede a eliminar las columnas innecesarias y cambiar el nombre a la última columna

In [25]: # Procederemos a eliminar columnas innecesarias
df = df.drop("id", axis=1)
# Cambiamos el nombre a la última columna
df = df.rename(columns={"species": "label"})
df.head()
Out[25]:
  sepal_length  sepal_width  petal_length  petal_width  label
0         5.1         3.5         1.4         0.2  iris-setosa
1         4.9         3.0         1.4         0.2  iris-setosa
2         4.7         3.2         1.3         0.2  iris-setosa
3         4.6         3.1         1.5         0.2  iris-setosa
4         5.0         3.6         1.4         0.2  iris-setosa
```

Ilustración 7 Código de los arboles de decisiones

## Bosques aleatorios

En el bosque aleatorio se hizo lo mismo que en el árbol, primeramente se da una explicación de lo que es y en que consiste.

Luego se explica cada función implementada en el mismo.

### Teoria

Jupyter Bosques\_Aleatorios Last Checkpoint: hace 24 minutos (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

### BOSQUES ALEATORIOS

Definición: Es un algoritmo de aprendizaje supervisado que, como ya se puede ver en su nombre, crea un bosque y lo hace de alguna manera aleatorio. Para decirlo en palabras simples: el bosque aleatorio crea múltiples árboles de decisión y los combina para obtener una predicción más precisa y estable. En general, mientras más árboles en el bosque se vean, más robusto es el bosque.

#### Se dividen en

#### Clasificación

Los valores de los conjuntos que se desean reproducir han de ser categorías. Siendo uno de los casos más habituales diferencias únicamente entre dos posibles categorías: verdadero o falso. Al igual que en los problemas de regresión es necesario identificar las variables independientes, la fórmula y los parámetros con los que se predecirá la variable dependiente. Solo que en este caso la variable dependiente es discreta.

#### Regresión

Los valores de los conjuntos que se desean reproducir han de tener un valor continuo. Siendo este el valor que se desea reproducir. Para ello se seleccionan las características con las que se realizarán las predicciones. A las que se denominan variables independientes. Una vez realizado esto se ha de definir una fórmula matemática y calcular los parámetros de modo que al introducir las características se obtenga el valor deseado. Esta se conoce como variable dependiente.

#### Funcionamiento

- Crea un bosque y lo hace de alguna manera aleatorio
- Crea múltiples árboles de decisión y los combina para obtener una predicción más precisa y estable
- Al crecer los árboles busca la mejor característica entre un subconjunto aleatorio de características
- Se puede hacer que los árboles sean más aleatorios, usando umbrales aleatorios para cada función

#### Diferencia entre los árboles de decisión y los bosques

Ilustración 8 teoria de los Bosques aleatorios

### Código

Jupyter Bosques\_Aleatorios Last Checkpoint: hace 25 minutos (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

```
111 [42]: A = df_train.drop('clase', axis=1)
112         y = df_train['clase']

In [26]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

In [27]: clf = RandomForestClassifier(n_estimators=4, max_depth=4)
         clf.fit(X_train, y_train)

         y_predict = clf.predict(X_test)
         metrics.accuracy_score(y_test, y_predict)

Out[27]: 0.725

In [28]: with open("wine-tree.dot", "w") as f:
         f = tree.export_graphviz(clf.estimators_[1], out_file=f, impurity=True, rounded=True, filled=True)

         os.environ["PATH"] += os.pathsep + 'C:/Program Files (x86)/Graphviz2.38/bin/'
         check_call(['dot', '-Tpng', 'wine-tree.dot', '-o', 'wine-tree.png'])

         PImage("wine-tree.png")

Out[28]:
```

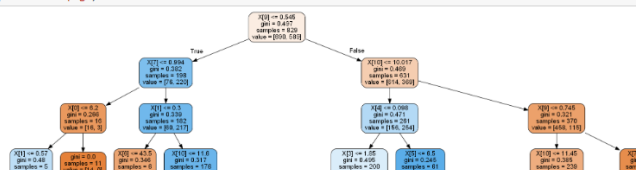


Ilustración 9 Código de los Bosques aleatorios

## **Conclusión**

El árbol de decisión es uno de los pocos algoritmos de Machine Learning que produce una comprensión comprensible de como el algoritmo toma de decisiones. Y así sabiendo como implementarlo en Jupyter Notebook podemos decir que se hace mucho más fácil la aplicación del mismo.

Los bosques aleatorios es un gran algoritmo para entrenar temprano en el proceso de desarrollo del modelo para ver cómo se desempeña y es difícil de construir un mal modelo con este algoritmo debido a su simplicidad.

Es una excelente opción, si necesita desarrollar un modelo en un corto periodo de tiempo, además de eso, proporciona un buen indicador de la importancia que asigna sus características.

## Información consultada

- [Wikipedia](#)
- [Bookdown](#)
- [Aprende Machine Learning arboles](#)
- [Aprende Machine Learning bosques](#)

## Puntaje

<b>Nombre</b>	<b>Puntaje</b>
Abel Enrique Consuegra Zuniga	10%
Heidy Nohemy Lemus Ramos	10%
Manuel Enrique Aguilar Reyes	10%
Ramón Humberto Ventura Rodríguez	10%
Will Alejandro Lara Mejía	10%