

Aplicación del aprendizaje automático con árboles de decisión al estudio de las variables del modelo de indicadores de gestión de las universidades públicas

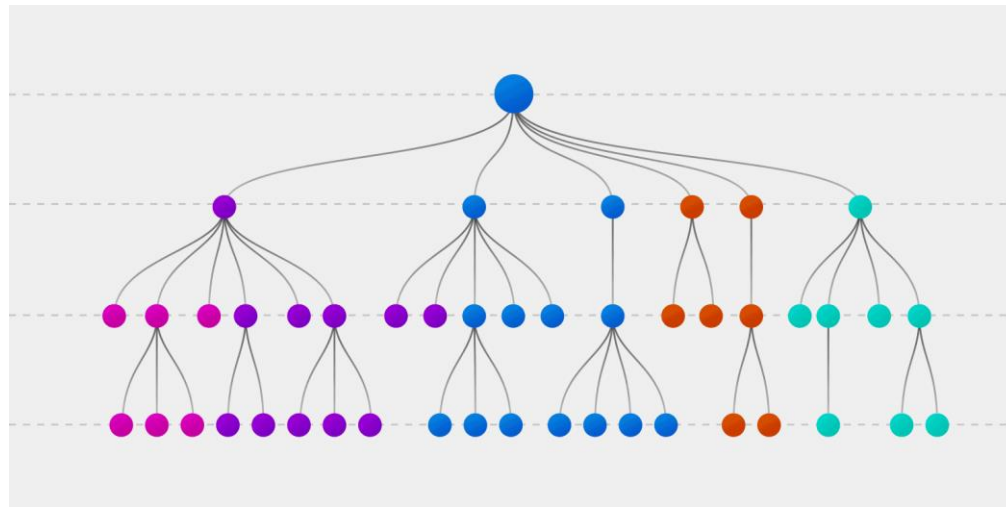
MODELADO MATEMÁTICO- PRESENTACIÓN 2



ALAN RENÉ CERDA
SERRATO 1864878
JORGE ANTONIO CRUZ
LERMA 1867341

MODELO ÁRBOLES DE DECISIÓN

- Un árbol de decisión es un modelo de predicción utilizado en diversos ámbitos que van desde la inteligencia artificial hasta la Economía.
- Dado un conjunto de datos se fabrican diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.



VENTAJAS

- Los árboles de decisión son simples de entender y de interpretar.
- Si el diagrama no es demasiado grande, puede visualizarse de un vistazo y eso facilita mucho el entendimiento del proceso.
- No requiere que prepares unos datos excesivamente complejos.
- Es válido tanto para variables cuantitativas como para cualitativas.
- Se pueden agregar nuevas opciones a los árboles existentes.
- Son fácilmente combinables con otras herramientas de tomas de decisiones.

DESVENTAJAS

- Son inestables: cualquier pequeño cambio en los datos de entrada puede suponer un árbol de decisión completamente diferente.
- No se puede garantizar que el árbol generado sea el óptimo.
- Hay conceptos que no son fácilmente aprendidos pues los árboles de decisión no son capaces de expresarlos con facilidad.
- Los principiantes crean árboles sesgados.
- Un árbol de decisión puede llegar a ser demasiado complejo con facilidad, perdiendo su utilidad

Aplicación del modelo en la minería de datos.

- De las herramientas disponibles en minería de datos tenemos las de descubrimiento del conocimiento, donde se encuentran los Algoritmos de árboles de decisión como una de las principales herramientas predictivas que ayudan a determinar los atributos de mayor incidencia en una decisión positiva o negativa entregándolos en un árbol y mostrando las reglas que guían al destinatario sobre cómo obtener esta decisión. Es un algoritmo TDIDT cuya heurística principal es buscar el mejor atributo para ubicarlo en la raíz del árbol y para esto utiliza un estadístico llamado mayor ganancia de información el cual está expresado como la diferencia de las siguientes entropías según el teorema de SHANON

La entropía de conjunto dada por la siguiente expresión:

$$H(S) = -p_p \log_2 p_p - p_n \log_2 p_n$$

$$H(S, A) \equiv \sum_{v \in \text{Valores}(A)} \frac{|S_v|}{|S|} H(S_v)$$

$$G(S, A) \equiv H(S) - H(S, A)$$

- La entropía ponderada definida como el valor esperado de la entropía del conjunto cuando se particiona de acuerdo a un atributo A, en particular. Esta entropía se calcula de acuerdo a la siguiente expresión: Con las dos expresiones anteriores se puede ahora obtener el estadístico de ganancia de información para cada atributo A de la información mediante la siguiente expresión: Es utilizando el mayor valor de esta ganancia que se selecciona el atributo que debe ir ubicado en la raíz del árbol.
- Como segundo elemento de la Heurística básica tenemos la redefinición del conjunto base para cada rama del árbol que corresponde a cada valor del atributo seleccionado en el paso anterior. Este proceso debe repetirse recursivamente para cada nuevo nodo generado en el árbol. Cuando un nodo no tiene descendientes distintos entonces se convierte en una rama que contiene la decisión a tomar.

Resumen y Problemática

- El estudio de las variables del modelo de indicadores de gestión del Sistema de Universidades Públicas (SUE) desde el punto de vista institucional, se convierte en un estudio de carácter prioritario, ya que actualmente el Ministerio de Educación asigna los recursos a distribuirse entre las universidades públicas con base en estos indicadores.
- En este estudio se analizan las variables que inciden dentro del modelo de indicadores de gestión y se propone una metodología general basada en árboles de decisión, específicamente el método de algoritmo que ayudará a elaborar un modelo matemático para determinar los ajustes necesarios que debe hacer la Universidad Tecnológica de Pereira, con el fin de mejorar sus índices de gestión.

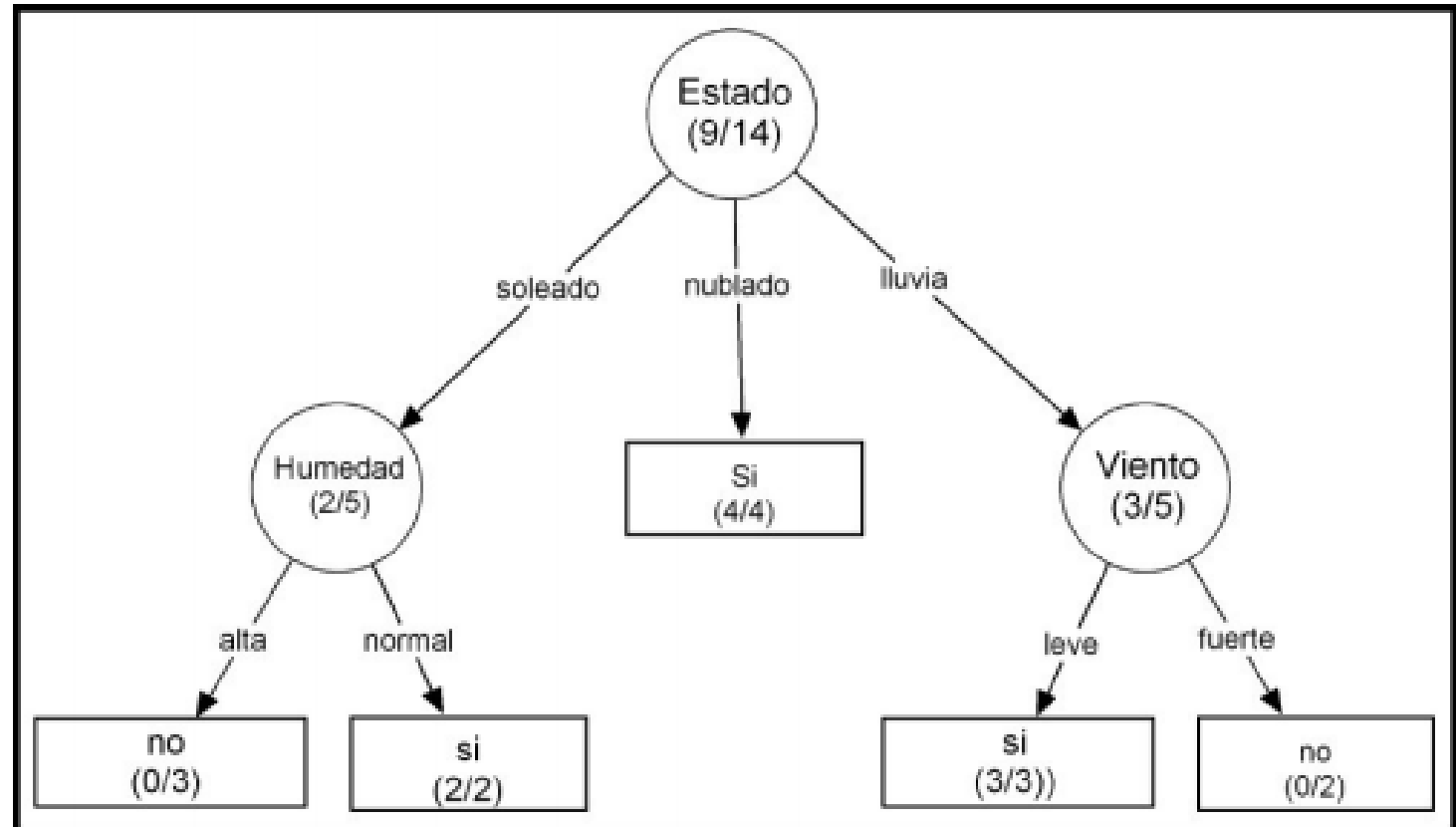
Ejemplo (Decisión sobre jugar Tenis)

- La teoría de este algoritmo sobre la decisión de ir o no a jugar tenis condicionado al valor de otras variables:

Información Base:

Estado	Temp	Humedad	Viento	Juego tenis
Soleado	Alta	Alta	Leve	No
Soleado	Alta	Alta	Fuerte	No
Nublado	Alta	Alta	Leve	Si
Lluvia	Suave	Alta	Leve	Si
Lluvia	Baja	Normal	Leve	Si
Lluvia	Baja	Normal	Fuerte	No
Nublado	Baja	Normal	Fuerte	Si
Soleado	Suave	Alta	Leve	No
Soleado	Baja	Normal	Leve	Si
Lluvia	Suave	Normal	Leve	Si
Soleado	Suave	Normal	Fuerte	Si
Nublado	Suave	Alta	Fuerte	Si
Nublado	Alta	Normal	Leve	Si
Lluvia	Suave	Alta	Fuerte	No

Después de aplicar el problema a los árboles de decisión, queda así:



Y las reglas producidas por el algoritmo:

La persona juega tenis si:

Estado=Soleado y Humedad=normal o

Estado=Nublado o

Estado=lluvia y Viento=Leve

En la etapa de evaluación y aplicación se observa la facilidad de entendimiento, por parte del usuario destinatario, tanto del árbol como de las reglas de decisión lo cual es tomado como referente para seleccionar el patrón hallado.

Bibliografía

- Problema de Google Académico:

Santa Chávez, Jhon Jairo; Veloza Mora, Juan De Jesús; Arias Montoya, Reinel Aplicación del aprendizaje automático con árboles de decisión al estudio de las variables del modelo de indicadores de gestión de las universidades públicas Scientia Et Technica, vol. 18, núm. 4, diciembre-, 2013, pp. 725-731 Universidad Tecnológica de Pereira Pereira, Colombia

<https://www.redalyc.org/pdf/849/84929984020.pdf>

<https://interactivechaos.com/es/manual/tutorial-de-machine-learning/ventajas-y-desventajas>