Tarea E6

Alexander Zapata Galindo – Código 201425426

Write at least 300 words explaining the types of decision trees algorithms and applications. You can send your answer through this form: https://forms.gle/ZQKhgwzpVxHTSqp89. Due date: 13/07/2021 5:59 pm.

Un árbol de decisión es un modelo analítico que contiene diagramas de construcciones o reglas lógicas que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva para la resolución de un problema. Estos diagramas son estructuras formadas por ramas y nodos de distintos tipos: i) Los nodos internos representan cada una de las características o propiedades a considerar para tomar una decisión; ii) Las ramas representan la decisión en función de una determinada condición; y iii) Los nodos finales representan el resultado de la decisión.

Esta técnica de toma de decisiones ha sido utilizada en diferentes campos y diversos ámbitos de la economía, política, biología, sociología, psicología, filosofía y ciencias de la computación y de los datos.

En términos generales, los árboles de decisión son algoritmos que siguen una secuencia lógica y estructurada de un proceso de decisión que permite evaluar un conjunto de alternativas de decisión (estrategias), teniendo en cuenta los eventos probabilísticos relevantes para la toma de la decisión y el comportamiento de cada alternativa con respecto a la variable de desempeño; el modelo selecciona aquella estrategia que produce el mejor valor esperado de la variable de desempeño. En esta vertiente son útiles software y programas especializados como DPL (<https://www.syncopation.com/decision-tree-software>), Canva (<https://www.canva.com/es_es/graficos/arboles-decision/>), Zingtree (<https://zingtree.com/>), SilverDecisions (<http://silverdecisions.pl/>), entre otros.

Específicamente en “machine learning”, los árboles de decisión son un algoritmo de aprendizaje supervisado que es usado para resolver problemas tanto de clasificación como de regresión, donde las variables de entrada y de salidas pueden también ser tanto continuas como categóricas. En esta vertiente, los árboles de decisión se clasifican en i) árboles de clasificación y ii) árboles de regresión.

De una parte, en los árboles de clasificación se busca entender el valor de una variable mediante la clasificación de la información en función de otras variables (tipo, pertenencia a un grupo, etc); la variable dependiente es categórica y el valor en el nodo terminal se reduce a la moda de las observaciones del conjunto de entrenamiento que han “caído” en esa rama. Por ejemplo, entender qué personas compran un determinado producto, clasificando entre clientes y no clientes, o qué marcas de portátiles compra cada persona mediante la clasificación entre las distintas marcas.

Por otra parte, en los árboles de regresión se intenta obtener el valor de una variable en función de otras variables que son independientes entre sí, la variable dependiente es continua y los valores de los nodos terminales se reducen a la media de las observaciones en cada rama. Por ejemplo, el precio de venta del terreno en función de variables como su localización, superficie, distancia a la playa, etc.

En “machine learning” son populares las diferentes librerías de software especializados como R (<https://www.r-project.org/>, con librerías como “rpart”, “party”,”randomForest”, “XGboost”, Phyton (<https://www.python.org/>, con librerías como “scikit-learn”, “XGboost”), Weka (<https://www.cs.waikato.ac.nz/>), Orange (<https://orangedatamining.com/>), entre otros.