GUÍA DE LECTURA VII (b): Árboles de decisión

1. ¿Qué es un árbol de decisión?
2. ¿Cuáles son los requerimientos para utilizar árboles de decisión?
3. ¿Cómo funciona ID3? ¿Cuáles son sus limitaciones?
4. ¿Por qué en el ejemplo de Play Tennis la variable Outlook accede a la raíz del árbol?
5. ¿Qué ocurre en ID3 con las variables con muchas etiquetas?
6. ¿Cuáles son las mejoras que se incorporan en C4.5?
7. ¿Cuál es la relación entre overfitting y la altura del árbol? ¿Cómo se resuelve?
8. ¿Qué ocurre si mi conjunto de datos está muy desbalanceado, es decir tiene 85% de los casos etiquetado como A y un 15% etiquetado como B?

Bibliografía sugerida:

* Mitchell, T. M. (1997). Machine learning. 1997. McGraw Hill, 45(37), 870-877. **Cap 3**. [[pdf](https://drive.google.com/file/d/0B_M_OiPKbozedWs2XzV6N2JPaDA/view?usp=sharing&resourcekey=0-c2Ojv34B1BqNi7SHlfbUlQ)]
* Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4). John Wiley & Sons. **Cap 8** [ [pdf](https://drive.google.com/file/d/0B5RFUdXfMb1iZXFhQlpqV014OUk/view?usp=sharing&resourcekey=0-2WUqBmWAWgz06vYAjSHLFw) ]