

Sílvia Benet, Marc Garrido

Tipologia i cicle de vida de les dades, pràctica 2.

URL github: <https://github.com/silviabenet/TipologiaCicleVidaDades>

Quina pregunta/problema pretén respondre?

El que pretenem respondre és a la pregunta de quina variable té més pes en la qualitat del vi i segons uns valors donats de les variables d'entrada volem saber de quina qualitat serà el vi resultant. Per tant aplicarem diferents algoritmes que ens facin una predicció de la variable qualitat.

Les dades contenen zeros o elements buits? Com gestionaries aquests casos?

Després de comprovar que les dades no contienien valors nuls hem comprovat els outliers. Pel que hem vist, hem trobat alguns valors outliers, pero semblaven valors dins d'un rang correcte, per tant no hem fet res amb ells. Un exemple és totalSulfureDioxide. Hem trobat dos valors outliers de la variable totalSulfureDioxide i hem vist que els dos registres tenien qualitat = 7.

A partir dels resultats obtinguts, quines són les conclusions?

L'atribut més important és clarament la quantitat d'alcohol, no només perquè és el primer node en particionar el camí sinó que més abaix també és crucial alhora de predir els vins de més qualitat.

Els altres atributs també importants veiem que serien els dels sulfats, volatilitat i potser el pH.

Els resultats permeten respondre al problema?

En efecte, la última visualització amb decision trees, podem veure de forma clara el camí de decisions per arribar a extreure'n la puntuació de qualitat d'un vi.

Contribucions	Firma
Investigació prèvia	Marc Garrido, Sílvia Benet
Redacció de les respostes	Marc Garrido, Sílvia Benet
Desenvolupament del codi	Marc Garrido, Sílvia Benet