میدانیم شاخص جینی شکست از شاخص های جینی افرازهای به وجود آمده تشکیل میشود

و تقریبا به صورت مشابه بهره اطلاعاتی هم از انتروپی های افراز های تشکیل شده.

پس میتوان برای تفاوت ها شاخص جینی و انتروپی را باهم مقایسه نمود.

به طور کلی میتوان گفت با انتخاب هرکدام تفاوت شدیدی در پروسه ما ایجاد نمیشود بجز در موارد خاص حتی در کتاب Tan هم جمله زیر را ذکر شده:

"Impurity measure are quite consistent with each other... Indeed, the strategy used to prune the tree has a greater impact on the final tree than the choice of impurity measure."

بااینحال در بعضی حالات این انتخاب تفاوت هایی ایجاد خواهد کرد.

همچنین بدلیل وجود لگاریتم در فرمول انتروپی و درنتیجه بهره اطلاعاتی، این روش کمی آرام تر و از لحاظ عملیاتی کمی سنگین تر است.

میتوان به ناخالصی جینی به عنوان تقریب درجه اول بهره اطلاعاتی نگاه کرد و درواقع بهره اطلاعاتی دقیقتر است.

شاخص جینی در افراز های بزرگترعملکرد بهتری دارد و اجرای آن بسیار آسان است.

ولی چون انتروپی از لگاریتم احتمالات استفاده میکند ساپورت بسیار بهتری برای افرازهای کوچک تر و همچنین توابع احتمالاتی exp و لاپلاس ایجاد میکند.

شاخص جینی در الگوریتم CART استفاده میشود درحالیکه بهره اطلاعاتی در ID3, C4.5 استفاده میشود.

زمان پردازش داده های رسته ای، جینی به صورت اسپلیت های باینری و به شکل success و failure نتیجه میدهد اما تفاوت انتروپی را بعد و قبل از اسپلیت میسنجد.

منابع:

https://www.analyticssteps.com/blogs/what-gini-index-and-information-gain-decision-trees

https://analyticsindiamag.com/gini-impurity-vs-information-gain-vs-chi-square-methods-for-decision-tree-split/#:~:text=Information%20gain%20is%20calculated%20by,of%20each%20class%20from%20one.

https://datascience.stackexchange.com/questions/10228/when-should-i-use-gini-impurity-as-opposed-to-information-gain-entropy