Information Gain (*IG*) is also referred to as mutual information, representing the expected reduction in *H* (*where entropy, H, is the measure of disorder*) produced by partitioning the dataset according to a given attribute. The *IG* for an outcome Y from an attribute X represents the expected decrease in entropy of Y conditioned on X. The more information the factor provides regarding the landslide distribution, the higher the *IG* value [1]. The specific formula is as follows:

$$IG(Y,X) = H(Y) - H(Y|X) \tag{S1}$$

where $H(Y)$ represents the entropy of Y, and $H(Y|X)$ represents the entropy of Y given X.

$H(Y)$ can be written as:

$$H(Y) = -\sum_{i\epsilon 1}^{k} p_i \, log_2 \, p_i \tag{S2}$$

where p_i represents the proportion of instances belonging to class i in the data.

$H(Y|X)$ can be written as:

$$H(Y|X) = \sum_{j\epsilon X}^{j} p_j \, H(Y|X=j) \tag{S3}$$

where p_j represents the probability that attribute X takes on value j in the data, thus:

$$H(Y|X=j) = \sum_{i\epsilon 1}^{k} p_{(i|j)} \, log_2 \, p_{(i|j)} \tag{S4}$$

We take the distance to fault of all landslides as an example to calculate the IG.

IG(landslide, distance to fault)=H(landslides)-H(landslide, distance to fault)

		Probability		
		Total area	Landslide area	Nonlandslide area
Distance to fault	0-10	4	1	3
	10-20	4	3	1
	20-30	8	2	6
	>30	16	12	4
Total		32		

H(landslides)= -($\frac{18}{32} log_2 \frac{18}{32}$)- -($\frac{14}{32} log_2 \frac{14}{32}$)= 0.99 (S5)

H(landslides, distance to fault)

=*p*(0-10)**H*($\frac{1}{4}, \frac{3}{4}$)+*p*(10-20)**H*($\frac{3}{4}, \frac{1}{4}$)+*p*(20-30)**H*($\frac{2}{8}, \frac{6}{8}$)+*p*(>30)**H*($\frac{12}{16}, \frac{4}{16}$)

$$=(\frac{4}{32})*(-(\frac{1}{4}log_2\frac{1}{4})-(\frac{3}{4}log_2\frac{3}{4}))+(\frac{4}{32})*(-(\frac{3}{4}log_2\frac{3}{4})-(\frac{1}{4}log_2\frac{1}{4}))+(\frac{8}{32})*(-(\frac{2}{8}log_2\frac{2}{8})-(\frac{6}{8}log_2\frac{6}{8}))+$$

$$(\frac{16}{32})*(-(\frac{12}{16}log_2\frac{12}{16})-(\frac{4}{16}log_2\frac{4}{16}))$$

$$=\frac{1}{8}*0.81+\frac{1}{8}*0.81+\frac{1}{4}*0.81+\frac{1}{2}*0.81=0.81 \quad \text{(S6)}$$

IG(landslide, distance to fault)=H(landslides)-H(landslide,distance to fault)=0.99-0.81=0.18 (S7)

Therefore, the IG for the distance to fault as a predictor for landslide occurrence in the above example is 0.18.

Reference

1. Fan, X.M.; Yunus, A.P.; Scaringi, G.; Catani, F.; Subramanian, S.S.; Xu, Q.; Huang, R.Q. Rapidly evolving controls of landslides after a strong earthquake and implications for hazard assessments. *Geophys. Res. Lett.* **2020**, doi:10.1029/2020GL090509.