Similarity and Distance Metrics

$$\text{Raw Similarity}(v_1, v_2) = \begin{cases} 1.0 & \text{if } v_1 = 0 \text{ and } v_2 = 0, \\ 0.0 & \text{if } v_1 = 0 \text{ or } v_2 = 0, \\ 1 - \min\left(\dfrac{|v_1 - v_2|}{\max(|v_1|, |v_2|)},\, 1\right) & \text{otherwise.} \end{cases}$$

$$\text{Bootstrap Normalized Similarity} = \begin{cases} 1.0 & \text{if } \sigma_{\text{bootstrap}} = 0 \text{ and } |v_{\text{pred}} - v_{\text{train}}| = 0, \\ 0.0 & \text{if } \sigma_{\text{bootstrap}} = 0 \text{ and } |v_{\text{pred}} - v_{\text{train}}| \neq 0, \\ \max\left(0,\, 1 - \min\left(\dfrac{|v_{\text{pred}} - v_{\text{train}}|}{\sigma_{\text{bootstrap}} \cdot 3},\, 1\right)\right) & \text{otherwise.} \end{cases}$$

where

v_{pred} is the predicted value, v_{train} is the training value, $\sigma_{\text{bootstrap}}$ is the standard deviation from bootstrap samples.

$$\text{Distance from Ideal} = \left|\text{Normalized Similarity} - 1.0\right|.$$

$$\text{Combined Distance} = \frac{\text{Distance from Ideal}_{\text{licit}} + \text{Distance from Ideal}_{\text{illicit}}}{2}.$$