Are the Same Features Always the Most Important?
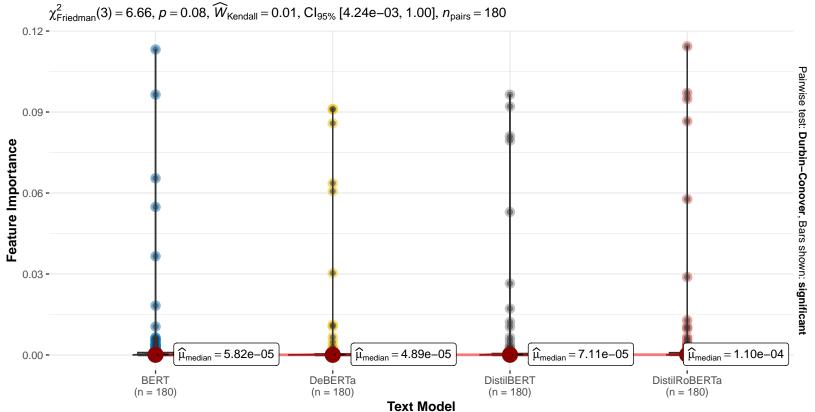**Feature Importance (SHAP) Compared, by Text Model**

$\chi^2_{\text{Friedman}}(3) = 6.66$, $p = 0.08$, $\widehat{W}_{\text{Kendall}} = 0.01$, $\text{CI}_{95\%}$ [4.24e−03, 1.00], $n_{\text{pairs}} = 180$

Feature Importance

Pairwise test: **Durbin-Conover**, Bars shown: **significant**

$\widehat{\mu}_{\text{median}} = 5.82\text{e}{-}05$

$\widehat{\mu}_{\text{median}} = 4.89\text{e}{-}05$

$\widehat{\mu}_{\text{median}} = 7.11\text{e}{-}05$

$\widehat{\mu}_{\text{median}} = 1.10\text{e}{-}04$

BERT
(n = 180)

DeBERTa
(n = 180)

DistilBERT
(n = 180)

DistilRoBERTa
(n = 180)

Text Model

Dataset: jigsaw