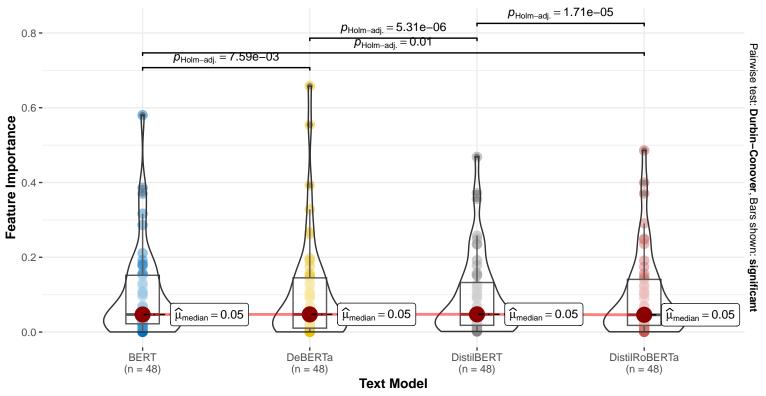**Are the Same Features Always the Most Important?**
Feature Importance (SHAP) Compared, by Text Model

$\chi^2_{\text{Friedman}}(3) = 29.30$, $p = 1.94e{-}06$, $\widehat{W}_{\text{Kendall}} = 0.20$, $CI_{95\%}$ [0.11, 1.00], $n_{\text{pairs}} = 48$

$p_{\text{Holm-adj.}} = 1.71e{-}05$

$p_{\text{Holm-adj.}} = 5.31e{-}06$

$p_{\text{Holm-adj.}} = 0.01$

$p_{\text{Holm-adj.}} = 7.59e{-}03$

Feature Importance

0.8 —
0.6 —
0.4 —
0.2 —
0.0 —

$\widehat{\mu}_{\text{median}} = 0.05$
$\widehat{\mu}_{\text{median}} = 0.05$
$\widehat{\mu}_{\text{median}} = 0.05$
$\widehat{\mu}_{\text{median}} = 0.05$

BERT
(n = 48)

DeBERTa
(n = 48)

DistilBERT
(n = 48)

DistilRoBERTa
(n = 48)

Text Model

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

Dataset: imdb_genre