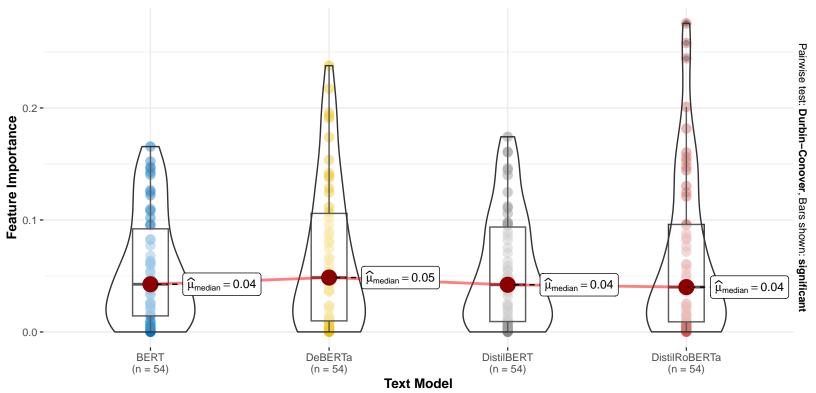**Are the Same Features Always the Most Important?**
Feature Importance (SHAP) Compared, by Text Model

$\chi^2_{\text{Friedman}}(3) = 1.40$, $p = 0.71$, $\widehat{W}_{\text{Kendall}} = 8.64\text{e}{-}03$, $\text{CI}_{95\%}$ [2.90e−03, 1.00], $n_{\text{pairs}} = 54$

Feature Importance

$\widehat{\mu}_{\text{median}} = 0.04$
$\widehat{\mu}_{\text{median}} = 0.05$
$\widehat{\mu}_{\text{median}} = 0.04$
$\widehat{\mu}_{\text{median}} = 0.04$

BERT
(n = 54)

DeBERTa
(n = 54)

DistilBERT
(n = 54)

DistilRoBERTa
(n = 54)

Text Model

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

Dataset: kick