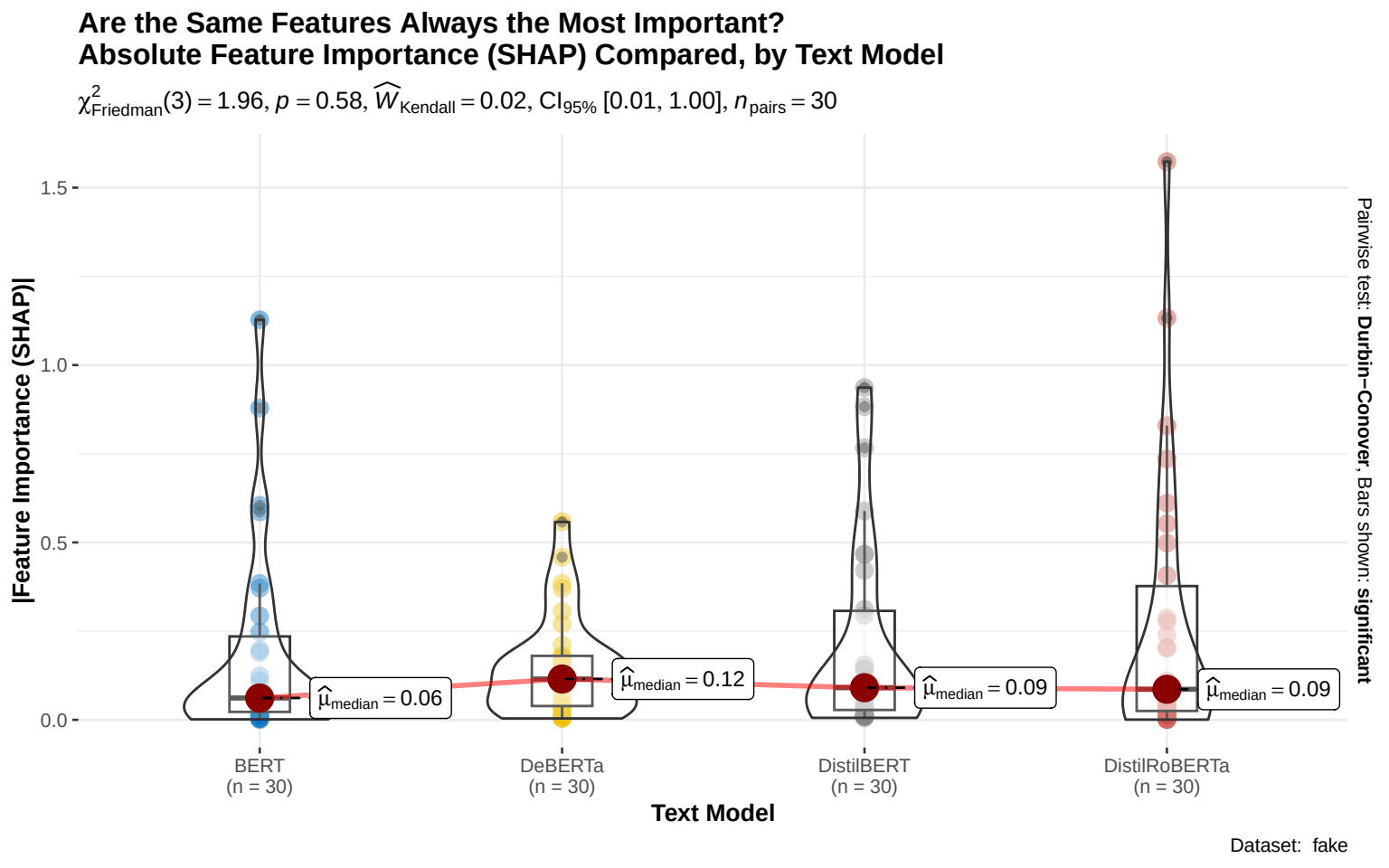
Are the Same Features Always the Most Important?
Absolute Feature Importance (SHAP) Compared, by Text Model
$\chi^2_{\text{Friedman}}(3) = 1.96$, $p = 0.58$, $\widehat{W}_{\text{Kendall}} = 0.02$, $\text{CI}_{95\%}$ [0.01, 1.00], $n_{\text{pairs}} = 30$
|Feature Importance (SHAP)|
1.5
1.0
0.5
0.0
$\widehat{\mu}_{\text{median}} = 0.06$
$\widehat{\mu}_{\text{median}} = 0.12$
$\widehat{\mu}_{\text{median}} = 0.09$
$\widehat{\mu}_{\text{median}} = 0.09$
BERT
(n = 30)
DeBERTa
(n = 30)
DistilBERT
(n = 30)
DistilRoBERTa
(n = 30)
Text Model
Pairwise test: Durbin−Conover, Bars shown: significant
Dataset: fake