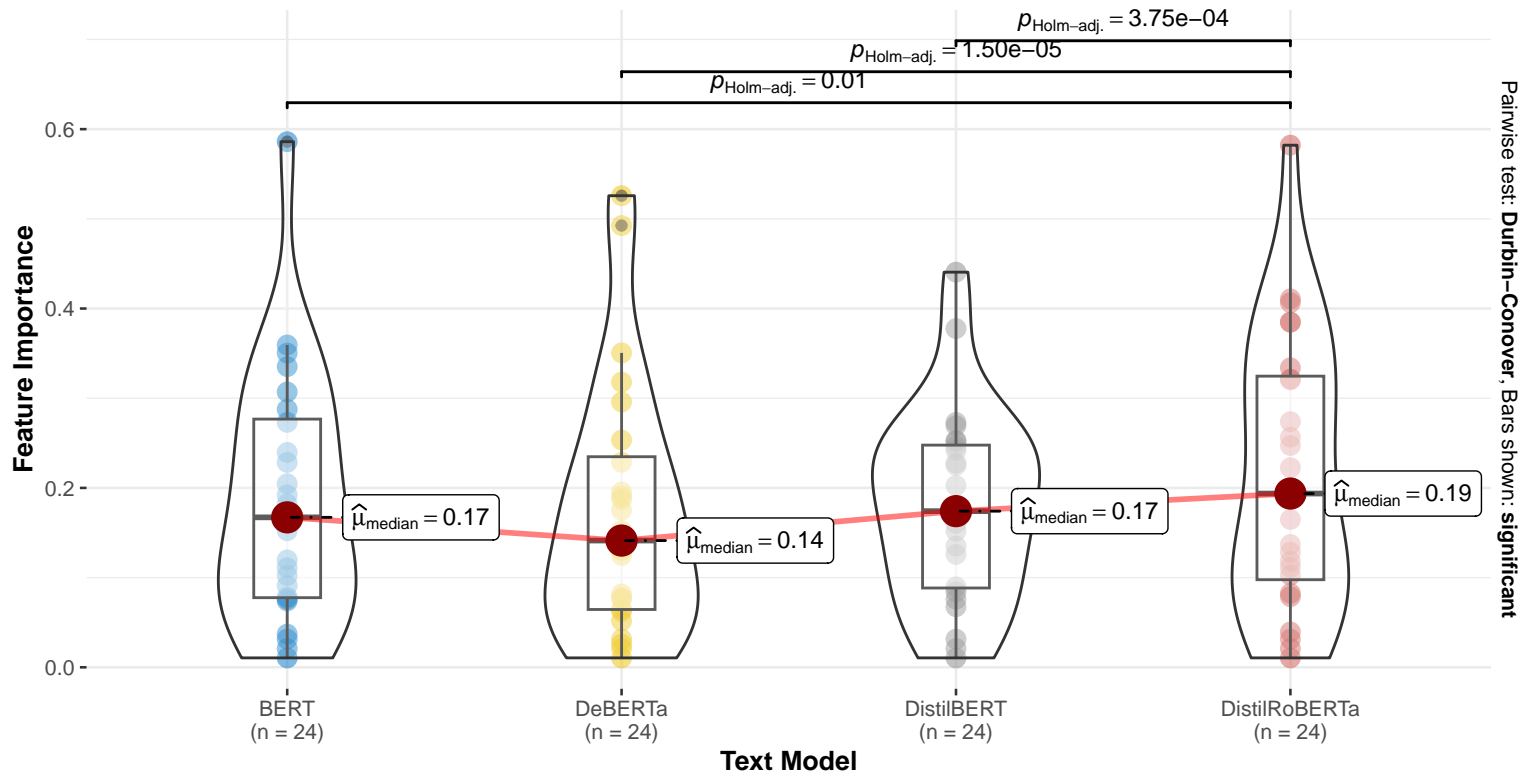**Are the Same Features Always the Most Important?**
Feature Importance (SHAP) Compared, by Text Model

$\chi^2_{\text{Friedman}}(3) = 21.81$, $p = 7.16\text{e}{-}05$, $\widehat{W}_{\text{Kendall}} = 0.30$, $CI_{95\%}$ [0.18, 1.00], $n_{\text{pairs}} = 24$

$p_{\text{Holm-adj.}} = 3.75\text{e}{-}04$

$p_{\text{Holm-adj.}} = 1.50\text{e}{-}05$

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

Feature Importance

0.6

0.4

0.2

0.0

$\widehat{\mu}_{\text{median}} = 0.17$

$\widehat{\mu}_{\text{median}} = 0.14$

$\widehat{\mu}_{\text{median}} = 0.17$

$\widehat{\mu}_{\text{median}} = 0.19$

BERT
(n = 24)

DeBERTa
(n = 24)

DistilBERT
(n = 24)

DistilRoBERTa
(n = 24)

Text Model

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

Dataset: salary *Stack models excluded