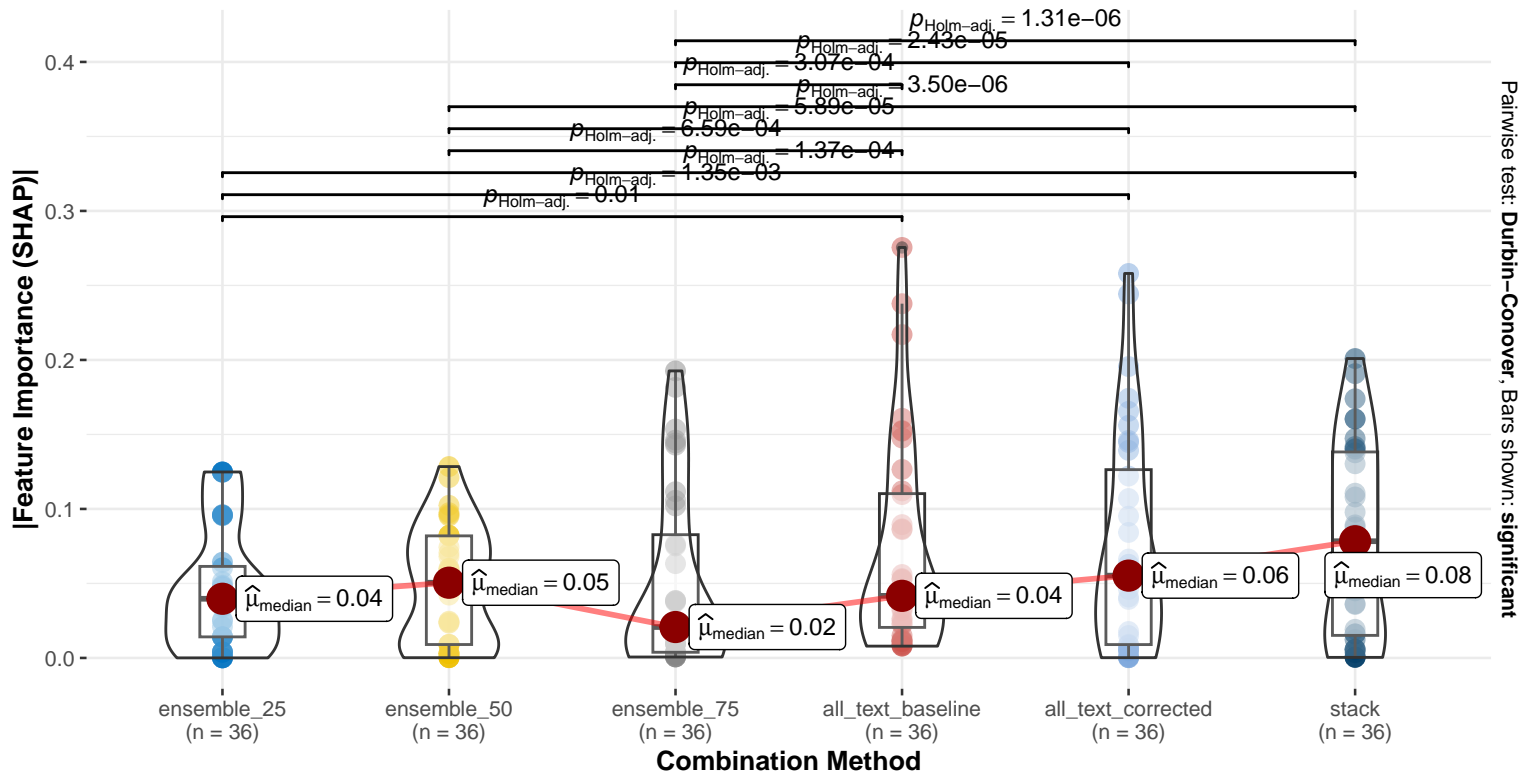**Are the Same Features Always the Most Important?**
Absolute Feature Importance (SHAP) Compared, by Combination Method

$\chi^2_{\text{Friedman}}(5) = 48.05$, $p = 3.47e{-}09$, $\widehat{W}_{\text{Kendall}} = 0.27$, $\text{CI}_{95\%}$ [0.20, 1.00], $n_{\text{pairs}} = 36$

$p_{\text{Holm-adj.}} = 1.31e{-}06$
$p_{\text{Holm-adj.}} = 2.43e{-}05$
$p_{\text{Holm-adj.}} = 3.07e{-}04$
$p_{\text{Holm-adj.}} = 3.50e{-}06$
$p_{\text{Holm-adj.}} = 5.89e{-}05$
$p_{\text{Holm-adj.}} = 6.59e{-}04$
$p_{\text{Holm-adj.}} = 1.37e{-}04$
$p_{\text{Holm-adj.}} = 1.35e{-}03$
$p_{\text{Holm-adj.}} = 0.01$

|Feature Importance (SHAP)|

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

$\widehat{\mu}_{\text{median}} = 0.04$
$\widehat{\mu}_{\text{median}} = 0.05$
$\widehat{\mu}_{\text{median}} = 0.02$
$\widehat{\mu}_{\text{median}} = 0.04$
$\widehat{\mu}_{\text{median}} = 0.06$
$\widehat{\mu}_{\text{median}} = 0.08$

| ensemble_25 (n = 36) | ensemble_50 (n = 36) | ensemble_75 (n = 36) | all_text_baseline (n = 36) | all_text_corrected (n = 36) | stack (n = 36) |

Combination Method

Dataset: kick