Head to head of feature sets

Mean balanced accuracy was calculated for each problem, with the highest declared winner