Tableshift: Voting 0.80 constant top all features 0.75 top causal features out-of-domain accuracy 09.0 29.0 29.0 29.0 top arguably causal features 0.55 0.50 0.50 0.55 0.60 0.65 0.70 0.75 0.80 in-domain accuracy