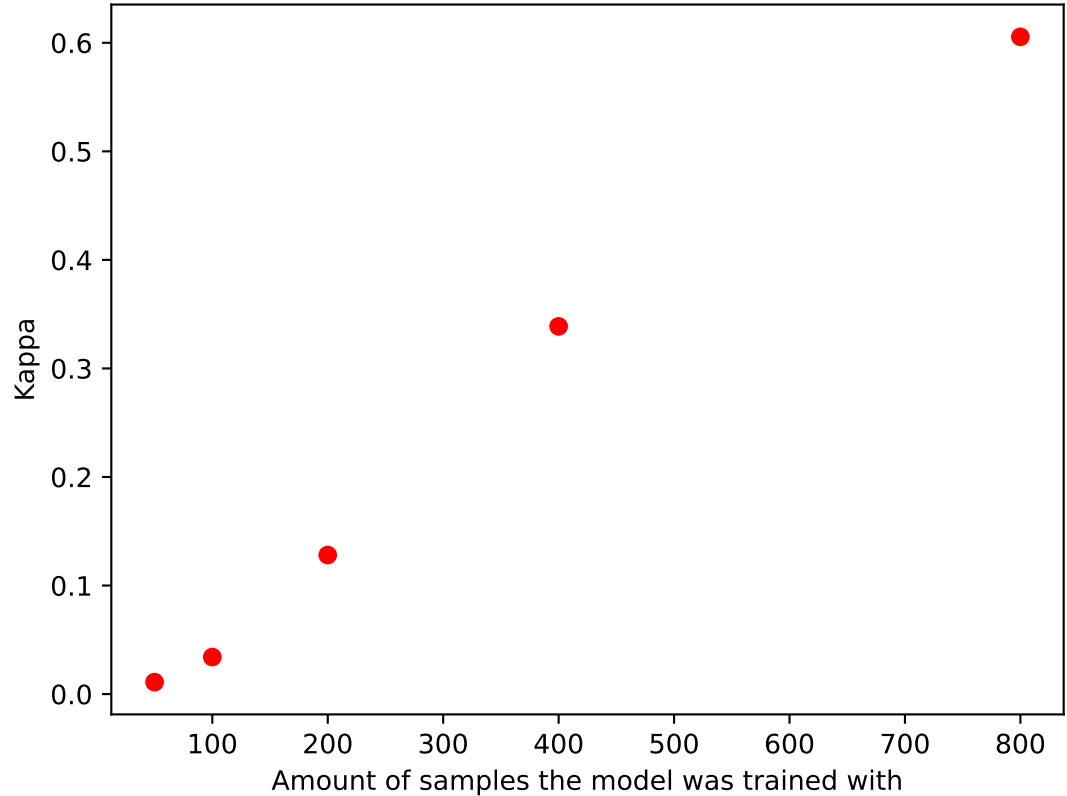XGB Model trained with samples created by text-davinci-003 rating answers created by human experts used for testing

Kappa

Amount of samples the model was trained with

Each value represents the agreement of exactly 392 ratings