Model performance over different training sets