Precision & Recall comparison on validation set (sorted by f1)