

Label-weighted average F1 score on test set

