We take jMeter as an example to give the reason why we only reported the *Recall* results.

Example: jMeter

As shown in Table 2, there are 260 classes and 14 (i.e., *TP*+*FN*=14) key classes. If we use a cutoff *k*=10 to filter out *non-key* classes, that is, the top-260∗10% (i.e., top-26) ranked classes are the *recommended as* key classes. Thus, (*TP*+*FP*) is equal to 260*10%=26, and (*TP* + *FN*) is equal to 14 (i.e., the number of true key classes in the *gold set*). Thus, in jMeter, (*TP*+*FP*=26) and (*TP*+*FN*=14) are fixed for different *unsupervised* approaches when *k*=10.

For example (**note that it is just an illustrative example; the numbers shown below are not the values obtained in our experiments**), when *k*=10, the *Precision, Recall,* and *F1* of Pride can be computed, respectively, as

$$Precision = \frac{TP}{TP+FP} = \frac{TP}{26}, Recall = \frac{TP}{TP+FN} = \frac{TP}{14}, F1 = \frac{2*\frac{TP}{26}*\frac{TP}{14}}{\frac{TP}{26}+\frac{TP}{14}}$$

The *Precision, Recall,* and F1 of iFit can be computed, respectively, as

$$Precision = \frac{TP}{TP+FP} = \frac{TP}{26}, Recall = \frac{TP}{TP+FN} = \frac{TP}{14}, \quad F1 = \frac{2*\frac{TP}{26}*\frac{TP}{14}}{\frac{TP}{26}+\frac{TP}{14}}$$

When comparing the *Precision* (or *Recall*) of Pride and iFit on the jMeter when *k*=10, their differences in *Precision* and *Recall* are mainly determined by *TP*. Thus, if Pride can identify two key classes in the gold set (i.e., *TP*=2) in the top-10% ranked classes, and iFit can identify five key classes in the gold set (i.e., *TP*=5), then the *Precision, Recall,* and *F1* of Pride can be computed, respectively, as

$$Precision = \frac{TP}{TP+FP} = \frac{2}{26}, Recall = \frac{TP}{TP+FN} = \frac{2}{14}, \quad F1 = \frac{2*\frac{2}{26}*\frac{2}{14}}{\frac{2}{26}+\frac{2}{14}} = 0.1$$

The *Precision, Recall,* and *F1* of iFit can be computed, respectively, as

$$Precision = \frac{TP}{TP+FP} = \frac{5}{26}, Recall = \frac{TP}{TP+FN} = \frac{5}{14}, \quad F1 = \frac{2*\frac{5}{26}*\frac{5}{14}}{\frac{5}{26}+\frac{5}{14}} = 0.25$$

iFit has a better *Recall* (when compared with the *Recall* of Pride) and also has a better *Precision* and *F1*. Thus, we believed that there is no need to report *Precision*, *Recall*, and *F1* at the same time in the manuscript. Due to the space limitation, we only show the results of *Recall* when the *nine* approaches are applied to the six subject systems.