Community Smell Occurrence Prediction on Multi-Granularity by Developer-Oriented Features and Process Metrics (Appendix)

1 Examples of Developers' Sentimental Comments

See Table 1.

2 Correlation Heatmap

See Fig.1.

3 Individual-Wide Prediction Performance (All Classifiers)

We demonstrate the results of all tested classifiers' performance in the three validation scenarios in Fig.2 and Fig.3. Meanwhile, Fig.2 and Fig.3 depict the mean (in solid lines), median (in dashed lines), as well as the rank of each classifier. Specifically, Fig.2 demonstrates the performance of classifiers using data balancing techniques, while the data for classifiers in Fig.3 are not balanced. The ranks of the classifiers are presented in the X-axis.

4 Individual-Wide Prediction Performance (Random Forest)

The performances of Random Forest (RF) are demonstrated in Figs.4(b), 5(b), and 6(b) for within-project, time-wise and cross-project prediction respectively. Moreover, we include the result of prediction using only the time-sensitive version of features proposed in our preliminary conference version [1], i.e., developer-oriented features incorporating experience and sentiment features in Figs.4(a), 5(a), and 6(a) to demonstrate the effectiveness of our proposed features.

The parametres of RF are displayed in Table 2.

5 Individual-Wide Prediction Feature Importance

Figs. 7-9 depict the mean (in solid lines), median (in dashed lines), as well as the rank of each feature's importance.

6 Community-Wide Prediction Performance

Fig.10 depicts the R^2 performance and their SK-ESD rankings of the models. Additionally, it depicts the mean (in solid lines), median (in dashed lines), and the rank of each model. We use R^2 as our primary comparator because it is a reliable metric which is easier to interpret [2], e.g., containing lower and upper bounds. Only models with mean performance greater than 0 are listed. The models using feature selection are marked with _FS in their suffixes. Multivariate approach such as TCN achieves negative R^2 in all projects, and thus it is excluded from Fig.10.

References

[1] Huang Z, Shao Z, Fan G, Gao J, Zhou Z, Yang K, Yang X. Predicting community smells' occurrence on individual developers by sentiments. In *Proc. 29th IEEE/ACM Int. Conference on Program Comprehension*, May 2021, pp. 230–241. DOI: 10.1109/ICPC52881.2021.00030.

[2] Chicco D, Warrens M J, Jurman G. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Comput. Sci.*, 2021, 7:Article No. e263. DOI: 10.7717/peerj-cs.623.

Table 1. Examples of Developers' Sentimental Comments

Feature	Project	Comment
High VAL	Camel	Cool, I will take care of this patch.
High ARO	Zookeeper	I will have time to work on it this week. I will try to post an initial version soon.
High DOM	Wicket	(from a code reviewer) Looks good. Feel free and commit.
High SAD	HBase	Sorry, that's my fault. That part doesn't work so I commented it out and forgotten about it.
High ANG	Hadoop HDFS	Who's going to takeout the trash if we half implement this? Lots of spec questions are still open!
High LOV	OFBiz	Thanks for the info. Have a good day, hopefully with not so much snow like here in Germany.
High JOY	Hadoop HDFS	Thanks for the patch! I learned some things on this one that surprised me a little bit. +1
High POS	HBase	Awesome! Thanks Todd.
High NEG	Hibernate ORM	This also fails!!! This bug is really annoying, please include the fix in the next release!
High POL	Hadoop Map/Reduce	It'd be nice to have the JobTracker object not be locked while accessing the HDFS.
Low POL	Hadoop HDFS	My committership does not make me smarter than you, but you can't force me commit blindly.
High IND	Cassandra	This was implemented as part of CASSANDRA-827.
High IMP	Cassandra	Apply this *after* the patch for CASSANDRA-963.
High CON	Hive	Not sure what happened (in a batch task). I would re-run. But there is quite a queue at present.
High SUB	Harmony	I suggest we rewrite the char methods to use ICU.
High MOD	Hadoop Common	This error is caused by Java 1.7 framework.
Low MOD	HBase	Does a test need to be written for this? It's not clear how easy that would be.

Table 2. Parametres of the Random Forest Models

Prediction Classes	Validation	n_estimators	max_depth	Sampling
Bottleneck	Within-Project	120	9	ROS
	Time-Wise	120	9	SMOTE
	Cross-Project	70	9	RUS
Lone Wolf	Within-Project	80	9	ROS
	Time-Wise	170	1	SMOTE
	Cross-Project	170	9	RUS
Organizational Silo	Within-Project	80	9	ROS
	Time-Wise	170	1	SMOTE
	Cross-Project	60	9	ROS
Smelly Developer	Within-Project	80	9	ROS
	Time-Wise	110	5	SMOTE
	Cross-Project	190	13	None
Smelly Quitter	Within-Project	30	1	RUS
	Time-Wise	40	1	ROS
	Cross-Project	10	5	RUS

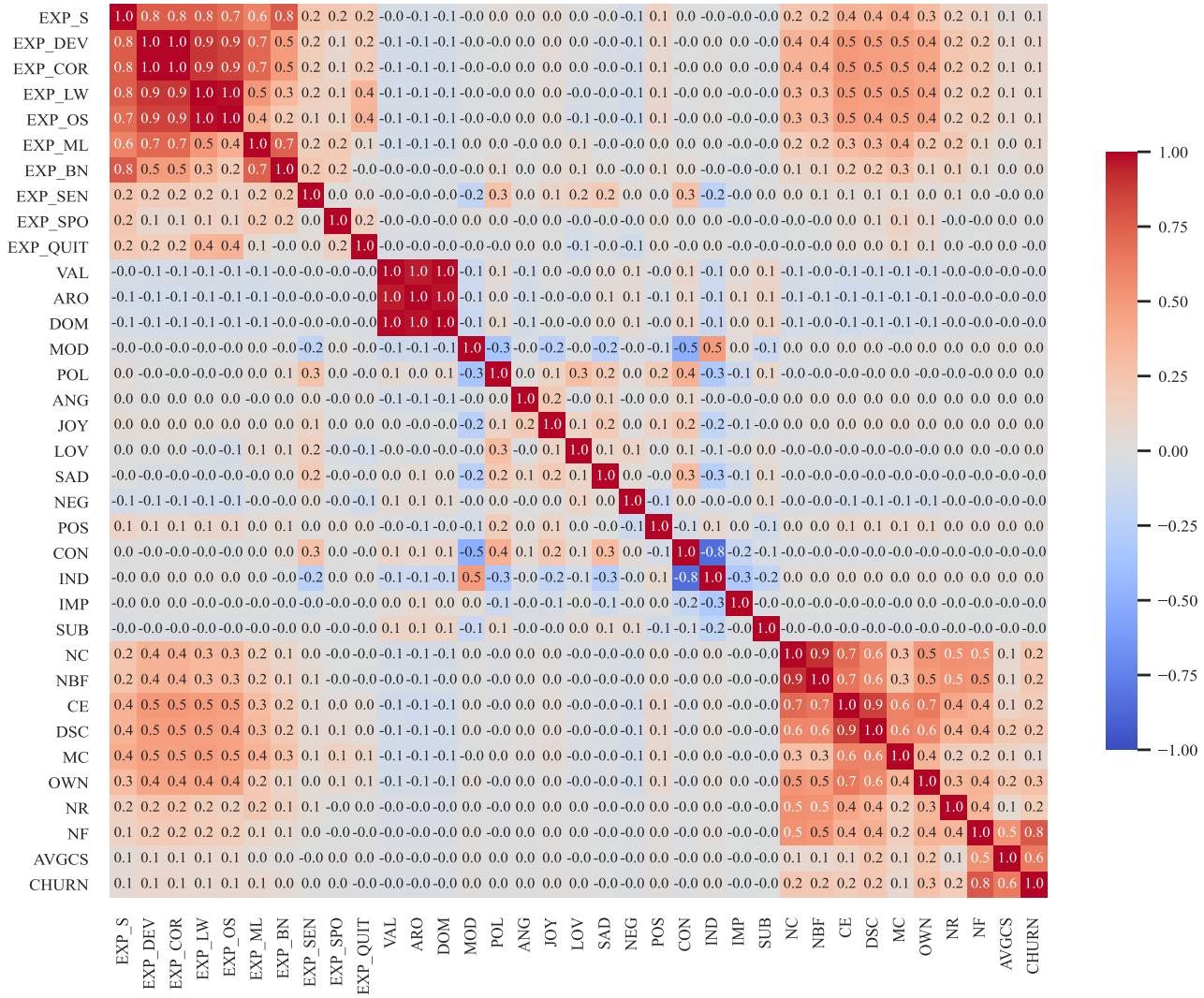

Fig.1. Correlation heat-map of features.

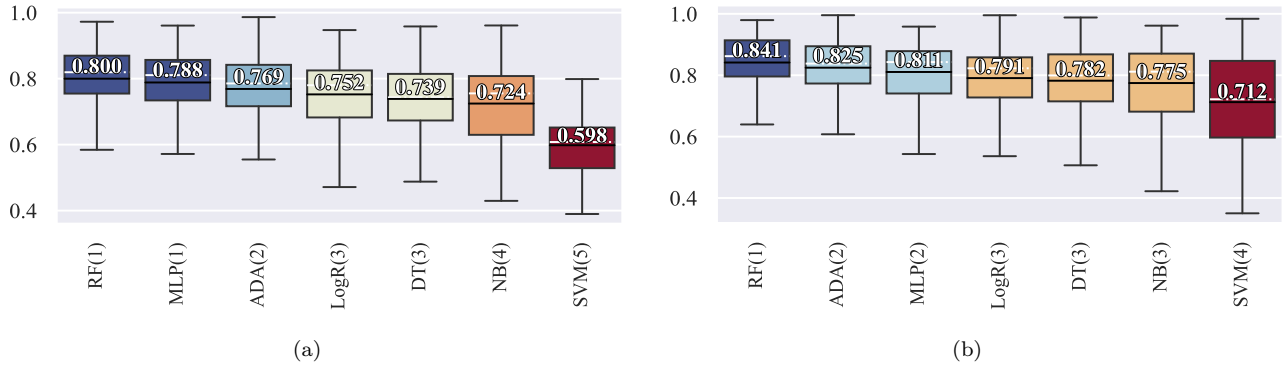

Fig.2. Individual-wide model performance with data balancing. (a) AUC-ROC. (b) *F*-Measure.

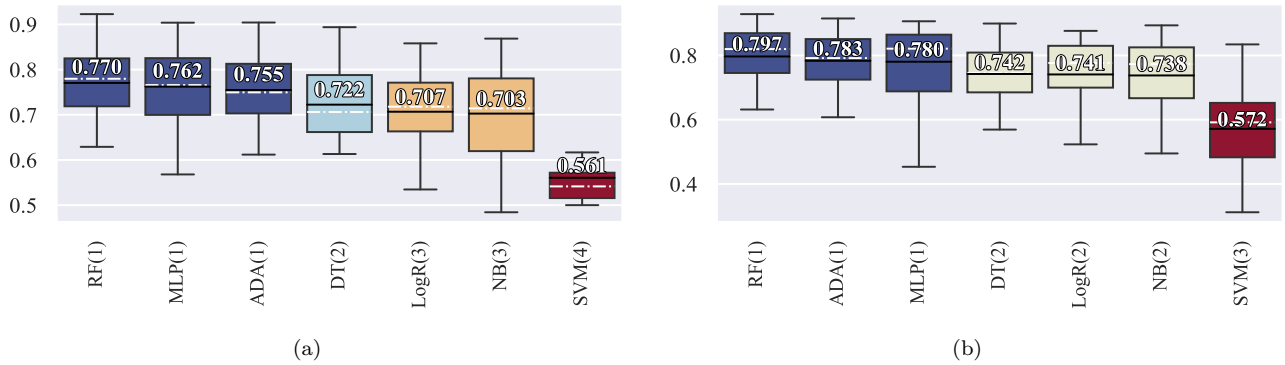

Fig.3. Individual-wide model performance without data balancing. (a) AUC-ROC. (b) *F*-Measure.

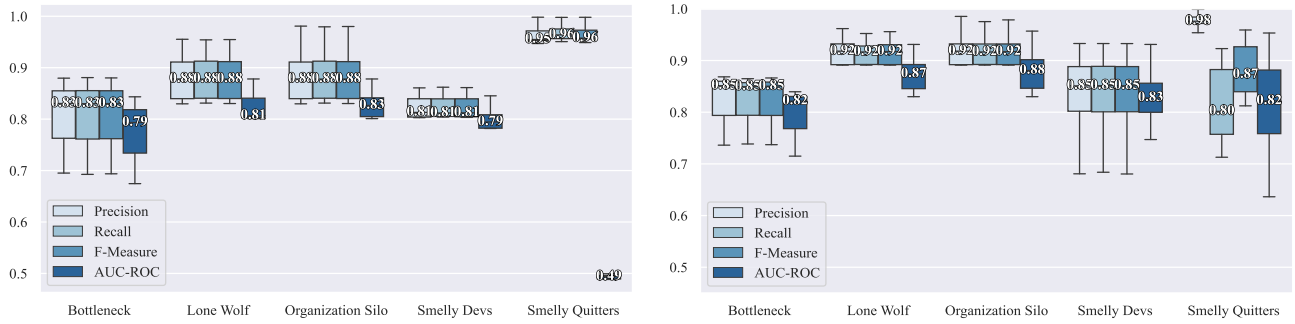

Fig.4. Within-project performance. (a) Only developer-oriented features. (b) Developer-oriented features and process metrics.

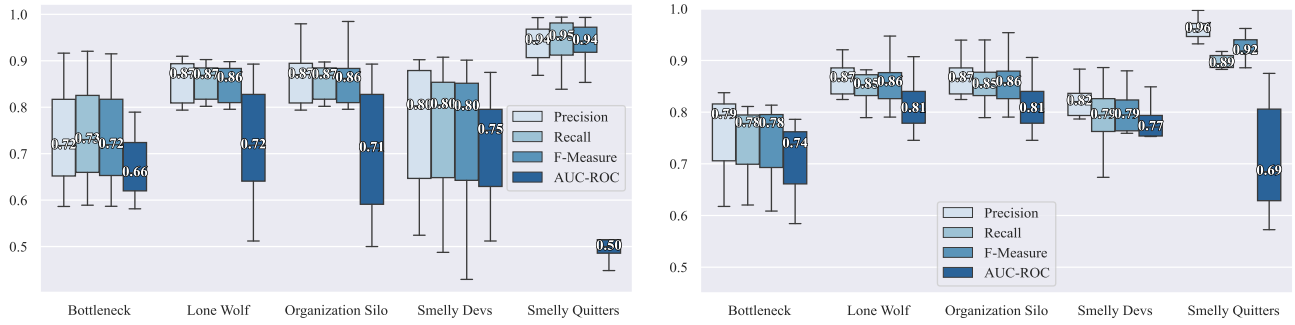

Fig.5. Time-wise performance. (a) Only developer-oriented features. (b) Developer-oriented features and process metrics.

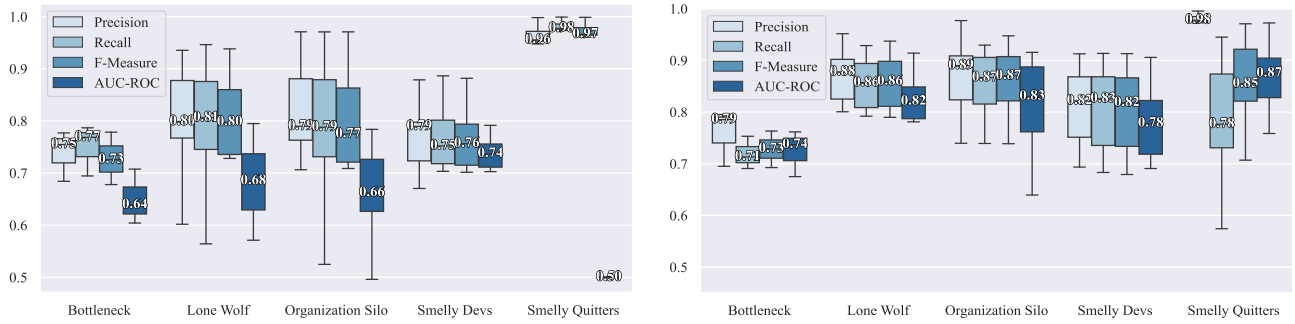

Fig.6. Cross-project performance. (a) Only developer-oriented features. (b) Developer-oriented features and process metrics.

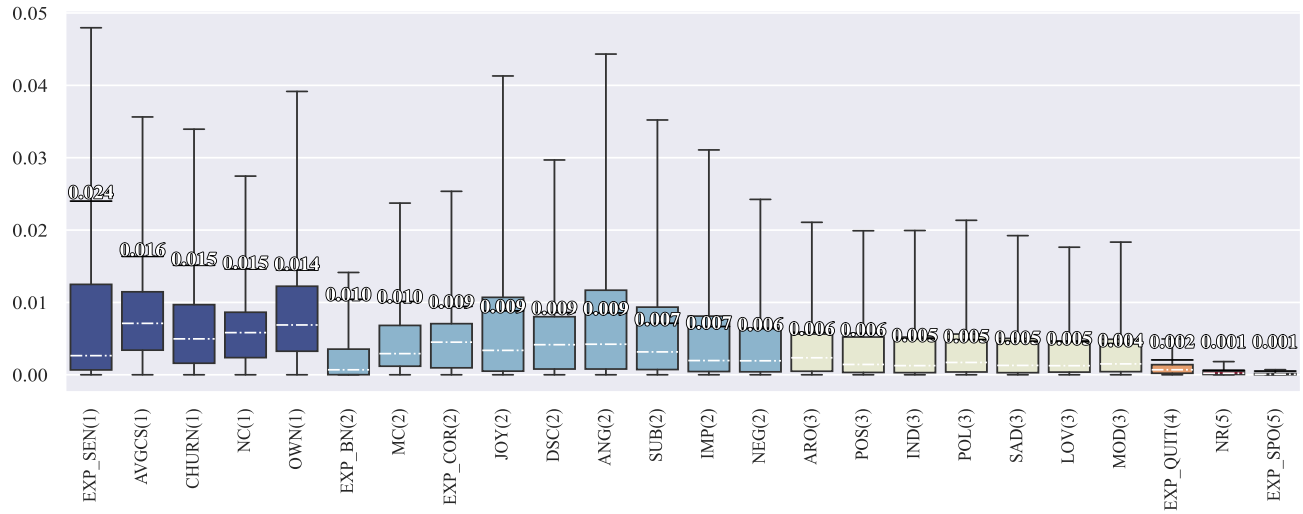

Fig.7. Within-project feature importance classified by SK-ESD. Outliers are hidden due to space limit.

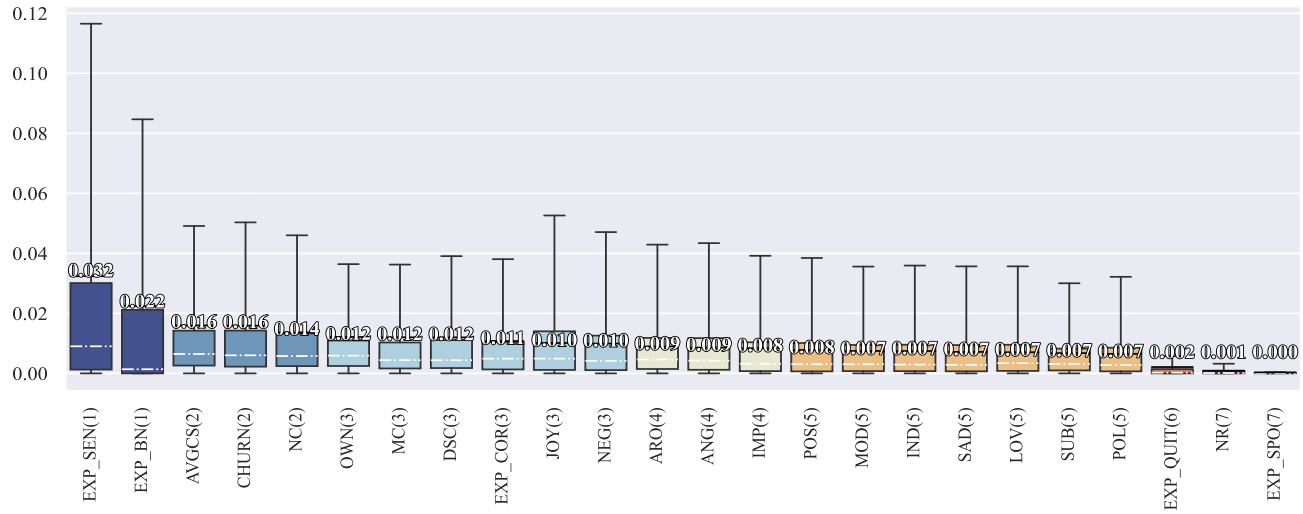

Fig.8. Time-wise feature importance classified by SK-ESD. Outliers are hidden due to space limit.

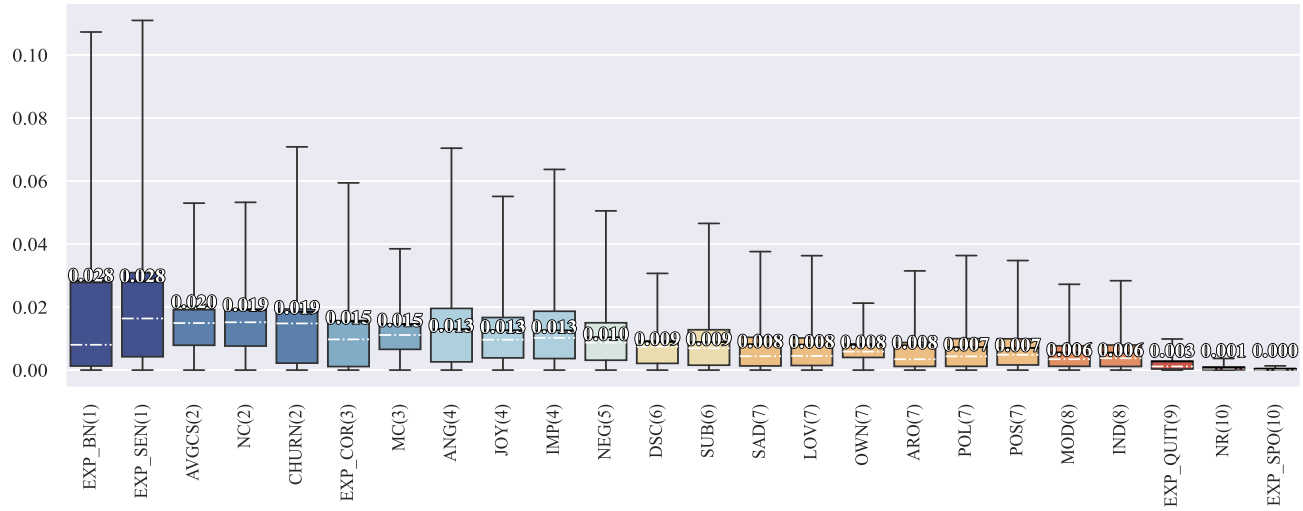

Fig.9. Cross-project feature importance classified by SK-ESD. Outliers are hidden due to space limit.

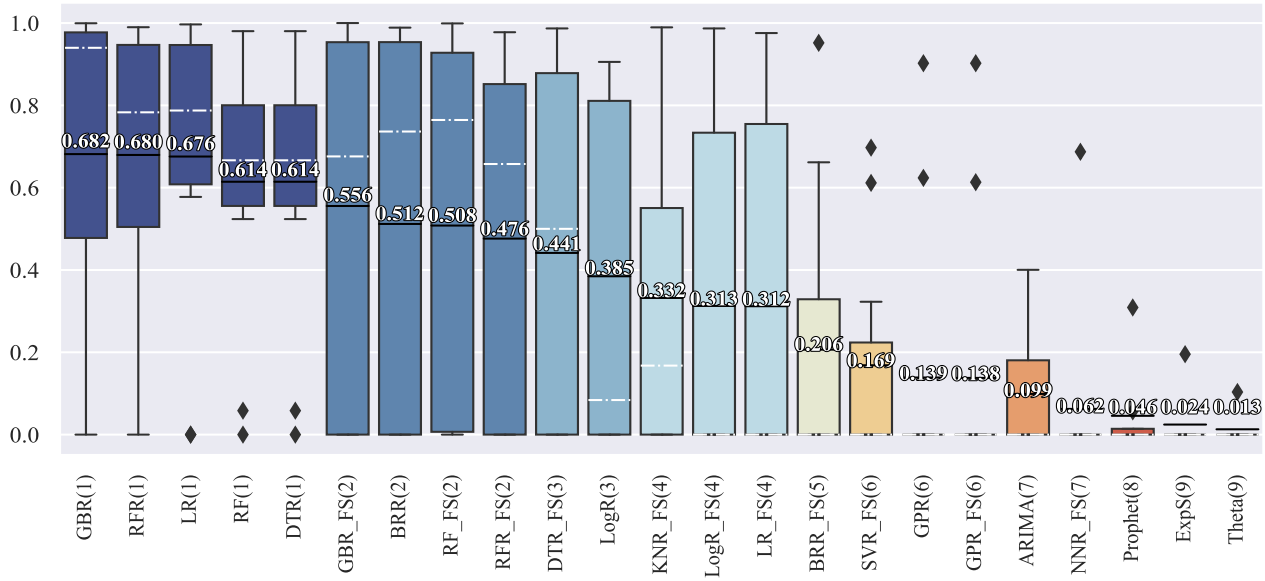

Fig.10. Community-wide model R^2 performance.