ICOIACT

by Rina Harimurti

Predicting Student's Psychomotor Domain on The Vocational Senior High School Using Linear Regression

R. Harimurti[1,2], Y. Yamasari[2], Ekohariadi[1], Munoto[1], I.G.P. Asto. B[1]
[1]Study Program of Vocational Education on Post Graduate, Universitas Negeri Surabaya Surabaya, Indonesia
[2]Department of Informatics, Universitas Negeri Surabaya, Surabaya, Indonesia
rinaharimurti@unesa.ac.id, yuniyamasari@unesa.ac.id, ekohariadi@unesa.ac.id, munoto@unesa.ac.id, asto@unesa.ac.id

***Abstract*— The educational data can be mined to produce the useful knowledge. This paper focuses on the educational data processing to predict student's psychomotor domain. Here, we apply linear regression method to do it. On process stage, we use 4 regularizations, namely: no regularization, ridge regression, lasso regression and elastic net regression. Furthermore, we exploit 2 sampling methods as the evaluation technique, for examples: cross-validation sampling and random sampling. The experimental result indicates that the best regularization on cross-validation and random sampling are an elastic net regression because this regularization achieves the lowest predicting error. On cross-validation, values of MSE, RMSE, and MAE are 40.079, 6.330 and 5.183, respectively. Additionally, for random sampling, respectively, values of MSE, RMSE, and MAE are 86.910, 8.428 and 6.511.**

Keywords—predicting; psychomotor; student; linear regression; educational data

I. INTRODUCTION

The psychomotor domain is a domain related to skill or ability to act after a person receives a particular learning experience. Psychomotor learning results are a continuation of cognitive and affective learning outcomes. If someone has shown a certain behavior or behavior in accordance with the purpose of the cognitive and affective aspects of the learning result can be said psychomotor learning results.

Skill learning is more effective when using learning by doing principles. It is published by Mills in 1977. Repeated skills will become a habit. There are some experts who put forward on how to assess psychomotor learning outcomes. Ryan [1] explains that skill learning outcomes can be measured through (1) direct observation and student behavioral assessment during the learning process of practice, (2) after learning, by giving test to learners to measure knowledge, and attitudes, (3) sometime after the learning is completed and later in the work environment.

Relating to psychomotor skills, psychomotor is associated with deliberate movements that are controlled by brain activity. Psychomotor is generally a skill that requires brain coordination with multiple muscles [2]. Motor skills not only demand the ability to sequence physical movement but also require mental/psychic activity (cognitive activity) to form an integrated movement of movement, so-called psychomotor ability. W.S. Winkel (1996).

For student achievement, in the process of learning to know the extent to which individuals carry out the learning process well and absorb the material taught well, it is necessary to measure the individual's learning ability. The measurement result is called learning achievement. In general, learning achievement is expressed in numbers or letters to compare with one criterion. Achievement or learning achievement which is a realization or a case of a person's potential skills or capacities.

Relating to student achievement, a teacher needs to collect data to do assessment process. It is called student data. For data student, there is the mining process which is done to get the useful information. For an example: clustering of student data which are collections of Massive Open Online Course is done by Yousef et. al [3]. It has a goal, that is to enhance the learning process. Yamasari et. al [4] mine a serious game-based student data by clustering method-FCM. It is addressed to more appropriate clustering of student's cognitive domain. In addition, Clustering can also be used to analyze the study length [5].

Furthermore, classification can be applied to the student data. The previous research has focused on it. They are Yuhana et. al. They do classification to predict the student's performance in math subject [6]. To improve student's motivation by Classification implementation to make the salary prediction system [7]. The building of the recommendation system based on prediction can also be used to help a student in the interesting course choosing [8]. This research also mines the educational data mining to predict student's psychomotor domain.

However, the prediction of previous research uses generally classification-based methods. Here, we apply linear regression for predicting of student's psychomotor domain. Furthermore, we focus on the error analyzing of this prediction. So, we measure the results by error metrics for examples: Mean Square Error, Root Mean Square Error and

978-1-5386-0954-5/18/$31.00 ©2018 IEEE

Figure 1. Display of Facebook media for problem-based learning model

Mean Absolute Error. In contrary, the previous research uses the classification metrics to evaluate the prediction result. They are accuracy, precision, recall, F1, etc.

Our paper consists of 4 sections, namely: Introduction is explained in section 1. For methods and material, they are presented in section 2. Next, the experimental result and discussion are explained in section 3. Finally, section 4 describes conclusion.

II. Methods and Material

This section comprises of two sub-sections. They are instructional media, instruments of data collecting process and methods.

A. Instructional Media

In this section, we describe the educational tool which is used to collect the student data. It is Facebook media for problem-based learning model presented in Figure 1. The process of instructional media building comprises of 2 stages. They are as follows:

Stage 1: the Facebook group creating. The learning process uses this group to conduct the problem-based learning model. The problem topic is posted by a teacher in this Facebook group.

Stage 2: learning process using Facebook media. This stage consists of 6 steps as follows:

Step 1: a menu of group list is used as discussion media.

Step 2: menu provided are group, member/ member register in group.

Step 3: menu for adding a photo, video, statement, attachment.

Step 4: column as the upload space or information sharing. It is used to upload the topic problem and to divide the teamwork with the learning instruction.

Step 5: comment column as an area for the student virtual interaction.

Step 6: the active chatting list and the inactive chatting list can be used for the Facebook user as the online discussion media who have become friend each other.

B. Instruments of Data Collecting Process

In the data collection process, we design and arrange several instruments as follows:

1) The syllabus is one of the curriculum development products that contains the outline of learning materials, learning activities, and learning design.
2) Learning Implementation Plan is a design implementation of learning process used by teachers in every one-on-one encounter, covering the basic competencies of the syllabus.
3) The module is a systematically organized teaching material consisting of aims to facilitate the teacher in the learning process.
4) Pretest evaluation as the introduction of a material or material and posttest aims to determine the competence that has been mastered by students as a result of the learning process
5) The questionnaire is a technique of data collection conducted by giving a set of questions or written statement to the respondent.

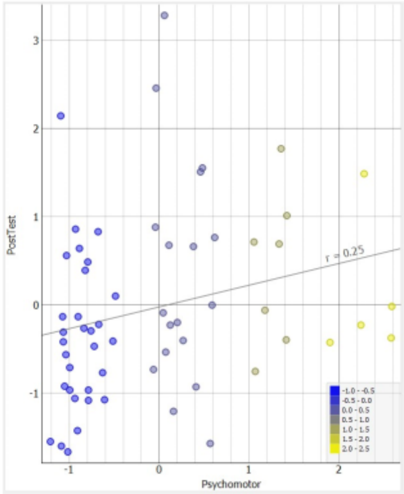

Figure 2. Scatter Plot for X=psychomotor and Y=PostTest

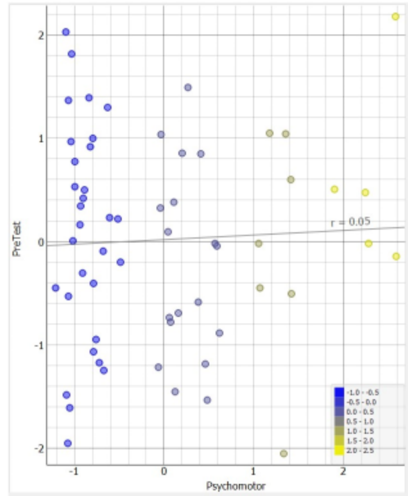

Figure 3. Scatter Plot for X=psychomotor and Y=PreTest

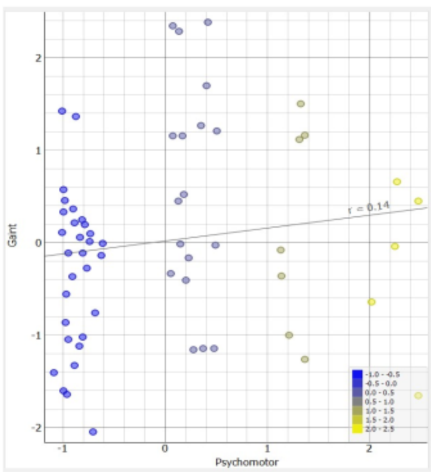

Figure 4. Scatter Plot for X=psychomotor and Y=Gain

C. Methods

In this section, methods and measures which are applied are explained, for example, Linear regression as the method to the modeling of data student which has many variables, namely: pretest, posttest, psychomotor and gain. In addition, metrics exploited are MAE, RMSE etc.

1) Simple Linear Regression

Simple Linear Regression is a statistical method that serves to test the extent of causality between variables predictor (X) to the response variable (Y). The model of the linear regression equation is as follows

$$Y = a + bX \qquad (1)$$

Where:

Y = response variable or dependent variable

X = predictor variable or independent variable

a = constant variable

b = regression coefficient

The eligibility requirements that must be fulfilled when going to use linear regression are as follows:

a. The number of samples used should be the same
b. The number of predictor variables (X) is 1
c. Residual value must be the normal distribution
d. There is a linear relationship between the predictor variable and the response variable.

2) Mean Absolute Error (MAE)

MAE (mean absolute error) is the average of the overall absolute error. Basically, MAE measures the average magnitude of error in a series of predictions, without considering its direction. This is used to test the absolute difference between the prediction and the actual observations in which all individual differences have the same weight. The formula of MAE is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |X_i - X| \qquad (2)$$

Where:

n = the number of error

$|X_i - X|$ = error absolute

TABLE I. PREDICTION RESULTS OF 4 REGULARIZATIONS

Student	Psychomotor	No_Regularization	Ridge_Regression	Lasso_Regression	Elastic_Net_Regression
1	82	80.275	80.277	80.276	80.318
2	88	81.683	81.695	81.692	81.972
3	82	76.684	76.69	76.689	76.842
4	94	80.004	79.997	79.99	79.835
5	75	79.142	79.143	79.144	79.161
9	94	80.563	80.564	80.563	80.582
…	…	…	…	…	…
58	75	81.083	81.085	81.086	81.127
59	88	79.76	79.749	79.743	79.485
60	88	80.461	80.461	80.462	80.476

3) Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) is the residual deviation standard (error prediction). The residue is a measure of how far from the point of regression data; RMSE is a measure of how this residue is distributed. So, it can be seen how the data is concentrated around the most appropriate line. The average square root error is used in climatology, forecasting, and regression analysis to verify the experimental results. The formula of RMSE is as follows:

The formula of RMSE is as follows:

$$\text{RMSE} = \sqrt{(f - \sigma)^2} \quad (3)$$

Where:
F = forecasts
σ = the observation value

III. RESULT AND DISCUSSION

In this section, we do an experiment for the various regularization related to linear regression. Visualization of linear regression is described in Figure 2. – Figure 4. Here, we try to paint the linear regression for axis x = Psychomotor feature and axis y = PostTest, PreTest, Gain, respectively. In these Figures, r (correlation) achieve the highest value for axis x = psychomotor feature and axis y = posttest. It is r = 0.25. For the linear regression with axis x = psychomotor feature and y = gain feature, it has r = 0.11.

The lowest value of r is shown on the linear regression with axis x = psychomotor feature and y = pretest feature. This means that the best correlation is regression linear created by axis x = psychomotor and axis y = posttest.

Furthermore, in this paper, we use many metrics to analyze whether the estimated linear regression model is good enough or in other words to measure how close the estimated regression line with the actual data. They are MSE, RMSE, MAE, and R2. These metrics are obtained by computing the prediction results. They are exploited of 4 regularizations on the psychomotor variable which is shown in Table I. Here, we do not display all record.

In our experiment, we set 2 sampling models to evaluate the prediction results, namely: Random Sampling and Cross-Validation. In addition, in every sampling, we apply 4 regularizations: no regularization, ridge regression, lasso regression and elastic net regression.

A. *Random Sampling*

We do the experiment by using many training set, namely: 5%, 10%, 20%, 25%, 30%, 33%, 40%, 50%, 60%, 60% Leave one out (LOO) and 60% Test On Train Data (TTD).

For this sampling, the result is described in Table II –V. All regularizations depict the lowest predicting error on the training set 60%(TTD). MSE = 33.73, RMSE = 5.808, MAE = 4.801 and R2 = 0.1 are achieved by no regularization, ridge regression and lasso regression. For elastic net regression, values of MSE, RMSE, MAE, and R2, respectively, are 33.755, 5.81, 4.8 and 0.099.

However, if the analyzing is based on an average of all experiments, we can show that the elastic net regression is the best regularization with the lowest values of MSE, RMSE, MAE, and R2 are 86.910, 8.428, 6.511 and -1.381, respectively.

Lastly, there are 3 regularizations which reaches the lowest predicting error, namely: no regularization, ridge regression and lasso regression on the specific training set. Moreover, the elastic net regression is the best regularization with the lowest values of all metrics.

B. *Cross-Validation Sampling*

In this sampling, we decide the experiment by using many folds, namely: 2, 3, 5, 10 and 20. The result of this experiment is presented in Table VI – Table IX.

For Table VI, it is obtained by no regularization. The best value of student's psychomotor domain prediction is

TABLE II. RANDOM SAMPLING-LINEAR REGRESSION_NO REGULARIZATION

Training set	MSE	RMSE	MAE	R2
5%	8425.346	91.79	48.594	-235.058
10%	984.26	31.373	19.206	-26.815
20%	68.678	8.287	6.33	-0.838
25%	53.836	7.337	5.632	-0.42
30%	48.068	6.933	5.297	-0.256
33%	47.067	6.861	5.245	-0.209
40%	43.923	6.627	5.148	-0.137
50%	38.671	6.219	5.123	-0.043
60%	35.888	5.991	5.048	0.023
60% (LOO)	38.737	6.224	5.152	-0.034
60% (TTD)	33.73	5.808	4.801	0.1
Average	892.564	16.677	10.507	-23.972

TABLE III. RANDOM SAMPLING-LINEAR REGRESSION_RIDGE REGRESSION

Training set	MSE	RMSE	MAE	R2
5%	594.802	24.389	15.823	-15.665
10%	254.36	15.949	11.493	-6.188
20%	68.886	8.3	6.329	-0.843
25%	53.965	7.346	5.629	-0.424
30%	48.09	6.935	5.298	-0.257
33%	47.088	6.862	5.246	-0.21
40%	43.92	6.627	5.147	-0.137
50%	38.639	6.216	5.121	-0.042
60%	38.723	6.223	5.151	-0.034
60% (LOO)	35.858	5.988	5.045	0.024
60% (TTD)	33.73	5.808	4.801	0.1
Average	114.369	9.149	6.826	-2.152

achieved on Fold 10 and fold 20. For fold 10, we can show that MSE and RMSE have the lowest value. The lowest MAE is reached on Fold 20. They indicate that in these folds, predicting error are the lowest value, that is MSE = 38.658, RMSE = 6.218 and MAE = 5.145. For Table VII, it is produced by ridge regression. The best value of student's psychomotor domain prediction is achieved on Fold 10 and fold 20.

The lowest MAE is reached on Fold 20. They indicate that in these folds, predicting error are the lowest value, that is MSE = 38.651, RMSE = 6.217 and MAE = 5.144.

Table VIII is generated by lasso regression. The best value of student's psychomotor domain prediction is

TABLE IV. RANDOM SAMPLING-LINEAR REGRESSION_LASSO REGRESSION

Training set	MSE	RMSE	MAE	R2
5%	6724.731	82.004	45.441	-187.411
10%	738.419	27.174	17.247	-19.868
20%	68.693	8.288	6.33	-0.838
25%	53.863	7.339	5.631	-0.421
30%	48.077	6.934	5.298	-0.257
33%	47.078	6.861	5.246	-0.21
40%	43.929	6.628	5.148	-0.137
50%	38.662	6.218	5.123	-0.042
60%	35.877	5.99	5.047	0.023
60% (LOO)	38.736	6.224	5.152	-0.034
60% (TTD)	33.73	5.808	4.801	0.1
Average	715.618	15.406	10.042	-19.009

TABLE V. RANDOM SAMPLING-LINEAR REGRESSION_ELASTIC NET REGRESSION

Training set	MSE	RMSE	MAE	R2
5%	381.616	19.535	14.348	-9.692
10%	164.771	12.836	9.539	-3.656
20%	69.564	8.341	6.38	-0.861
25%	54.609	7.39	5.619	-0.441
30%	48.257	6.947	5.304	-0.261
33%	47.251	6.874	5.252	-0.214
40%	43.916	6.627	5.144	-0.137
50%	38.343	6.192	5.101	-0.034
60%	35.51	5.959	4.999	0.033
60% (LOO)	38.421	6.198	5.13	-0.026
60% (TTD)	33.755	5.81	4.8	0.099
Average	86.910	8.428	6.511	-1.381

achieved on Fold 10 and fold 20. For fold 10, we can show that MSE and RMSE have the lowest value. The lowest MAE is reached on Fold 20. They indicate that in these folds, predicting error are the lowest value, that is MSE = 38.659, RMSE = 6.218 and MAE = 5.145.

Table IX is obtained with elastic net regression. The best value of student's psychomotor domain prediction is achieved on Fold 10. Here, we can show that MSE, RMSE, and MAE have the lowest value, namely: 38.539, 6.208 and 5.126, respectively.

However, based on an average of all experiments, we can show that the elastic net regression is the best regularization with the lowest values of MSE, RMSE,

TABLE VI. CROSS VALIDATION-LINEAR REGRESSION_NO REGULARIZATION

Fold	MSE	RMSE	MAE	R2
2	41.164	6.416	5.223	-0.099
3	42.6	6.527	5.293	-0.137
5	40.144	6.336	5.217	-0.072
10	**38.658**	**6.218**	5.151	-0.032
20	38.904	6.237	**5.145**	-0.039
Average	40.294	6.347	5.206	-0.076

TABLE VII. CROSS VALIDATION-LINEAR REGRESSION_RIDGE REGRESSION

Fold	MSE	RMSE	MAE	R2
2	41.126	6.413	5.22	-0.098
3	42.5	6.526	5.292	-0.137
5	40.135	6.335	5.216	-0.071
10	**38.651**	**6.217**	5.15	-0.032
20	38.893	6.236	**5.144**	-0.038
Average	40.261	6.345	5.204	-0.075

TABLE VIII. CROSS VALIDATION-LINEAR REGRESSION_LASSO REGRESSION

Fold	MSE	RMSE	MAE	R2
2	41.151	6.415	5.221	-0.099
3	42.5	6.525	5.291	-0.137
5	40.146	6.336	5.217	-0.072
10	**38.659**	**6.218**	5.15	-0.032
20	38.904	6.237	**5.145**	-0.039
Average	40.272	6.346	5.205	-0.076

TABLE IX. CROSS VALIDATION-LINEAR REGRESSION_ ELASTIC NET REGRESSION

Fold	MSE	RMSE	MAE	R2
2	40.79	6.387	5.185	-0.089
3	42.4	6.513	5.275	-0.133
5	39.998	6.324	5.2	-0.068
10	**38.539**	**6.208**	**5.126**	-0.029
20	38.667	6.218	5.127	-0.032
Average	40.079	6.330	5.183	-0.070

MAE, and R2, respectively, are 40.079, 6.330, 5.183 and -0.070.

Finally, for this sampling, the best regularization is an elastic net regression because it achieves the lowest predicting error if it is compared to the other regularizations. Furthermore, R2 have the minus value. It maybe happens because the number of data is too little.

IV. CONCLUSION

Prediction of student's psychomotor domain by using linear regression reaches the optimal value on elastic net regularization for random sampling and cross-validation sampling.

Furthermore, the evaluation results of cross-validation are better than random sampling for all metrics, namely: MSE, RMSE, and MAE based on average for all results.

ACKNOWLEDGMENT

The author would like to thank Study Program of Vocational Education on Post Graduate, Universitas Negeri Surabaya for its support on this research.

REFERENCES

[1] D. Kennedy, A. Hyland, and N. Ryan, "Implementing Bologna in your institution C 3.4-1 Writing and Using Learning Outcomes: a Practical Guide."

[2] R. N. SINGER, "4 – Motor Skills and Learning Strategies," in *Learning Strategies*, 1978, pp. 79–106.

[3] A. M. F. Yousef, M. A. Chatti, M. Wosnitza, and U. Schroeder, "A Cluster Analysis of MOOC Stakeholder Perspectives," *RUSC. Univ. Knowl. Soc. J.*, vol. 12, no. 1, p. 74, Jan. 2015.

[4] Y. Yamasari, S. M. S. Nugroho, I. N. Sukajaya, and M. H. Purnomo, "Features extraction to improve performance of clustering process on student achievement," in *2016 International Computer Science and Engineering Conference (ICSEC)*, 2016, pp. 1–5.

[5] P. Belsis, I. Chalaris, M. Chalaris, C. Skourlas, and A. Tsolakidis, "The Analysis of the Length of Studies in Higher Education based on Clustering and the Extraction of Association Rules," in *Procedia Social and Behavioral Sciences*, 2014, vol. 147, pp. 567–575.

[6] U. L. Yuhana, R. G. Mangowal, S. Rochimah, E. M. Yuniarno, and M. H. Purnomo, "Predicting Math performance of children with special needs based on serious game," in *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*, 2017, pp. 1–5.

[7] P. Khongchai and P. Songmuang, "Implement of salary prediction system to improve student motivation using data mining technique," in *2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, 2016, pp. 1–6.

[8] H. Bydovska and L. Popelinsky, "Predicting student performance in higher education," in *Proceedings - International Workshop on Database and Expert Systems Applications, DEXA*, 2013.

ICOIACT

ORIGINALITY REPORT

13%	10%	9%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ieeexplore.ieee.org Internet Source	1%
2	ijstr.org Internet Source	1%
3	Submitted to Universitas Negeri Surabaya The State University of Surabaya Student Paper	1%
4	repository.uki.ac.id Internet Source	1%
5	Ida Nugroho Saputro, Soenarto, Herminarto Sofyan. "Integrated Learning Factory Design on Future Learning", Walter de Gruyter GmbH, 2022 Publication	1%
6	id.123dok.com Internet Source	1%
7	journal2.uad.ac.id Internet Source	1%

8 Yuni Yamasari, Puput W. Rusimamto, Naim Rochmawati, Dwi F. Suyatno, Setya C. Wibawa, Supeno M. S. Nugroho, Mauridhi H. Purnomo. "Discretization method to optimize logistic regression on classification of student’s cognitive domain", MATEC Web of Conferences, 2018 <1%
Publication

9 S I Haryudo, A I Agung, R Firmansyah, W A Wibowo, Ekohariadi, Munoto. "The effect of the implementation of project based learning on learning activities of electrical engineering students", Journal of Physics: Conference Series, 2020 <1%
Publication

10 Submitted to Sriwijaya University <1%
Student Paper

11 Yuni Yamasari, Naim Rochmawati, Ricky E. Putra, Anita Qoiriah, Asmunin, Wiyli Yustanti. "Predicting the Students Performance using Regularization-based Linear Regression", 2021 Fourth International Conference on Vocational Education and Electrical Engineering (ICVEE), 2021 <1%
Publication

12 Submitted to University of Leicester <1%
Student Paper

13	eprints-phd.biblio.unitn.it Internet Source	<1 %
14	sinta.ristekbrin.go.id Internet Source	<1 %
15	sinta3.ristekdikti.go.id Internet Source	<1 %
16	eprints.utar.edu.my Internet Source	<1 %
17	link.springer.com Internet Source	<1 %
18	thinkmind.org Internet Source	<1 %
19	Submitted to Oxford Brookes University Student Paper	<1 %
20	docplayer.net Internet Source	<1 %
21	Submitted to UCSI University Student Paper	<1 %
22	Submitted to Universitas Brawijaya Student Paper	<1 %
23	mafiadoc.com Internet Source	<1 %
24	arxiv.org Internet Source	<1 %

25	eprints.uny.ac.id Internet Source	<1 %
26	Submitted to University of Leeds Student Paper	<1 %
27	journals.qou.edu Internet Source	<1 %
28	journals.uran.ua Internet Source	<1 %
29	www.matec-conferences.org Internet Source	<1 %
30	www.science.gov Internet Source	<1 %
31	Cheng Lei, Kin Fun Li. "Academic Performance Predictors", 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, 2015 Publication	<1 %
32	Veronika Zaporozhko, Denis Parfenov. "Automation of Controlling of Personalization of Learning in Cloud Educational Environment Based on Cluster Approach", 2018 International Russian Automation Conference (RusAutoCon), 2018 Publication	<1 %