



Improving Accuracy by applying Z-Score Normalization in Linear Regression and Polynomial Regression Model for Real Estate Data

Dimas Aryo Anggoro¹, Wiwit Supriyanti²

¹Informatics Department, Universitas Muhammadiyah Surakarta, Indonesia, dimas.a.anggoro@ums.ac.id

²Informatics Management Department, Politeknik Indonusa Surakarta, Indonesia, wiwitsupriyanti@poltekindonusa.ac.id

ABSTRACT

This research aimed to prove whether a normalization concept, i.e., z-score normalization was able to improve the accuracy result of algorithms for predicting house prices in Sindian District, New Taipei City, Taiwan. There were specific techniques in data science, e.g., linear regression and polynomial regression, that became the focus of this research. This study used several features that were implemented to model the house price. The features used included the transaction date, the house age, the distance to the nearest MRT, the number of nearby convenience stores, the latitude geographic coordinate, and the longitude geographic coordinate. The data were then preprocessed by splitting it into a training dataset (75%) and a testing dataset (25%) using a simple random sampling method. The subsequent step was to process the dataset with linear regression and polynomial regression model. Based on the result of the data processing, the optimum order befell to quadratic polynomial regression – maximum order was 2. This algorithm was then applied to normalized datasets and earned low scores of Mean Squared Error (MSE) and R-squared score, which were 7.044×10^{-7} and 0.989, respectively. It was concluded that this algorithm combination was the best-performed algorithm for predicting the real estate dataset in Sindian District, New Taipei City, Taiwan.

Key words: Real estate price, Linear regression, Polynomial regression, Z-score normalization

1. INTRODUCTION

Real estate has been one of the demanding market sectors in which most people are interested in purchasing. It cannot be hindered that owning a real estate is a necessity for people since the primary needs of people include foods, attires, and houses. Based on the research, the price of houses has been increasing. In general, the price of houses is also expected to increase in the future.

As an example, in 2008, the house price in Sindian District, New Taipei City, Taiwan, was around 500,000 TWD per ping – a local unit, 1 Ping = 3.3 meters squared. The house price was considered affordable if being compared to the current value of money. Whereas, in 2014, the house prices had experienced a significant increase until reaching approximately 850,000 TWD per ping [1]. The house prices had undergone a noticeable increase of about 70% in 6 years.

A further fact stated that in the future, it is projected that the size real estate data is going to be larger than the real estate data in the recent years. This might result in big data problem in which the dataset has large scale of volume, variety, and velocity – known as 3V. It is necessary to analyze big data to discover the knowledge behind the dataset [2]. Although the knowledge extraction was essential for big data issue, it was also applicable in the smaller size dataset collected in the recent years.

In this research, the knowledge of the dataset was extracted to predict the price of the real estate. The fluctuation of the house price was modeled to create a prediction model for real estate data. The real estate dataset is in the form of spatial and temporal data. In processing the dataset, data mining methods were required to be conducted. Data mining techniques vary depending on the dataset that was processed [3].

In finding out the pattern of data fluctuation, there were several methods proposed in previous works that had been conducted by some researchers. First, Ge and Wu [4] implemented linear regression models to detect the fluctuation of corn prices in China, starting from 2005 to 2016. The result showed that linear regression performed well – predicting the corn price with a relative error of 6.43%. Ge and Wu stated that multivariate linear regression produced more accurate results than the univariate linear regression model.

Additionally, the second work was conducted by [5]. In their research, higher-order multivariable polynomial regression was implemented. This model was claimed to improve the accuracy and simplicity of the model. Furthermore, they stated that linear regression lack of precision since the aspect of nonlinearity was being ignored. Their research focused on estimating human affective states using a polynomial regression model. The research informed that this algorithm performed well – resulting in correlation coefficients of 98% and 96% for predicting active valence and arousal, respectively.

This study examined whether the normalization process was able to improve the accuracy result of linear regression and polynomial regression for predicting the price of real estate in Sindian District, New Taipei City, Taiwan. Those regression algorithms were applied in different datasets to calculate the accuracy score of each regression technique. In computing the accuracy score of regression, Mean Squared Error (MSE) and R-squared (R²) tests were required to be conducted for evaluation [6], [7].

Hence, in this research, each regression model was combined, e.g., linear regression and polynomial regression model with z-score normalization, which was conducted in the data preprocessing phase. This technique was expected to improve the accuracy score of the model [8]. The output of the research was to prove whether the z-score normalization able to boost the accuracy score of the algorithm.

2. RESEARCH METHODS

The proposed method in solving the research problem was divided into several steps, as illustrated in Figure 1.

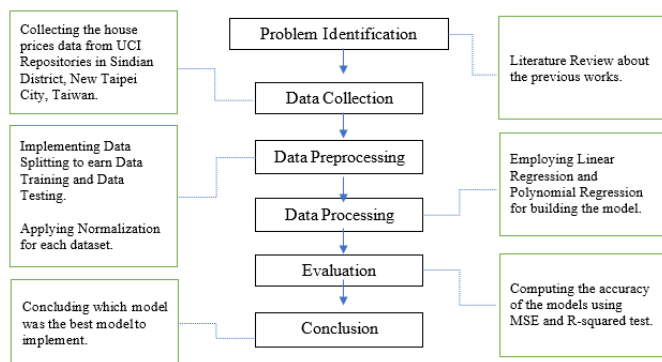


Figure 1: Research Stages

2.1. Problem Identification

This phase was the initial phase of the research method. In this phase, problems related to Linear regression and Polynomial regression produced satisfactory accuracy scores in predicting the previous works. Because of that, those regression methods were proposed in this research. Although the algorithms performed well in the previous research, the accuracy result for estimating real estate prices could still be improved. Hence, it was essential to combine each algorithm with another method, i.e., z-score normalization, to produce a new blended algorithm that was able to predict real estate prices more accurately.

2.2. Data Collection

The dataset was obtained from a secondary dataset from the University of California, Irvine data repository. The data were collected based on the real estate valuation in Sindian District, New Taipei City, Taiwan. The dataset contained 414 data points, also six independent variables (X1-X6), and one dependent variable (Y). The target variable (y) was the house price of the unit area (10,000 TWD per ping). Table 1 shows the details of features and targets.

Table 1: Table of features

Features	Details
X1	The transaction date (for example, 2016.500 = 2016 June, 2016.750 = 2016 September)
X2	The house age (year)
X3	The distance to the nearest MRT (meter)
X4	The number of convenience stores in the living circle on foot (integer)
X5	The geographic coordinate, latitude (degree)
X6	The geographic coordinate, longitude (degree)

2.3. Data Preprocessing

a. Data Splitting

In this stage, the dataset was split into two parts, which were training and testing datasets. The training dataset was used to produce a model; it still contained label data. Additionally, the testing dataset was used for the validation of the model – measuring the accuracy of the algorithm. In the testing dataset, the label data were removed and stored separately as true target values. Hence, the remaining value of the testing dataset was the input features. In this research, the dataset was split into 75% of data training (310 data points) and 25% of data testing (104 data points).

b. Z-score Normalization

Before creating a model from the training dataset, the dataset was required to be preprocessed. In this experiment, the preprocessing technique implemented was z-score normalization. In the research conducted by [8], this technique was claimed to be able to improve the accuracy of the model. This technique transformed the dataset into zero-mean and unit-variance distribution. The original dataset contained six features; each feature ranged differently. Therefore, the normalization process was conducted to transform the dataset into a specified range.

In this research, the performance of the two datasets was compared. One dataset was the original – without being processed with z-score normalization while the other was preprocessed with z-score normalization. The formula of z-score normalization is shown in Equation 1.

$$x'_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

where x'_i is normalized data, x_i is original data, μ is the average of data, and σ is the standard deviation of the data.

2.4. Data Processing

In processing the data, the algorithms employed were Linear Regression and Polynomial Regression. These algorithms were either combined or not with z-score normalization.

a. Linear Regression

Linear Regression was a regression algorithm that produced a linear model, which had order 1. This algorithm was implemented since it was suitable for prediction case for the continuous data as input [9]. In the research carried out by [10], the linear regression model processed numeric data; it modeled the input data to produce the numeric target. The technique performed fairly well with the accuracy of 0.6111. Linear regression was formulated in Equation 2.

$$y = \beta_0 + \sum_i \beta_i \cdot X_i \tag{2}$$

where y is the target, β_0 is intercept, β_i is weight, and X_i is data. Based on [11], however, the linear model was prone to be an under-fitted model, since the plotted model was in the form of a linear model with order 1. Nevertheless, it was still able to perform the general trend of the feature. Consequently, there was a high bias in the linear model; the test dataset was failed to be predicted accurately. Therefore, there was another regression algorithm proposed which was expected to perform a better accuracy score – less prone to be under-fitted.

b. Polynomial Regression

On the other hand, the second model that was discussed in this research was polynomial. This model performed in a more complex model than the linear model; and to some

extent, was expected to produce a better performance than the former model [12]. Nonetheless, the optimum number of degree that was applied in the polynomial model was required to be calculated before the execution of the algorithm. The reason was that a low degree was prone to underfit, while an overly high degree tended to overfit – including noise and outlier to form a model [13].

In the research conducted by [14], polynomial regression produced better performance with less variance value than the linear counterparts. Equation 3 shows the formula polynomial regression.

$$y = \beta_0 + \sum_p \beta_p \cdot X_j^p \tag{3}$$

where y is the target, β_0 is intercept, β_i is the feature's weight, X_j is data, and p is the order of the polynomial. Based on [15], the optimum degree needed to be computed prior to building a polynomial model. The arbitrary preference for optimum order could result in a less accurate model. The optimum order could be found when the model produced a minimum Sum of Square Errors (SSE) of the estimated responses. In this research, the proposed evaluation was MSE since it was directly proportional to SSE. The score of SSE was going to be decreased when the MSE score fell. The relation of SSE with MSE was formulated in Equation 4.

$$SSE = MSE (n - k) \tag{4}$$

where n is the number of data, and k is the number of samples.

2.5. Evaluation

After the models were generated, the real estate price was predicted by inserting the testing data into the model to obtain the predicted target values. If all predicted target values were the same as the true target, there would be zero error. In contrast, if there were some value differences between the predicted target and the true target, an error existed. In case the error was large, the model was considered as a weak model to predict the data, while the model was attributed as a strong model if the data were highly correlated with the true target value [16]. The evaluation phase was conducted by two different testing methods for regressions, including Mean Squared Error (MSE) and R-squared (R^2) tests.

3. RESULTS AND DISCUSSIONS

In this phase, two datasets were used. The first dataset was original, while the second one was the dataset with z-score normalization. The normalization process was claimed to produce a better model for prediction [8]. It was also strengthened by the implementation of the polynomial regression, which was believed to produce a better fit model than the linear counterpart. Therefore, this research hypothesized that Polynomial regression with z-score normalization generated the best model amongst all algorithm combinations.

As can be observed, the original dataset ranged differently on each different attribute. For example, the distance of the nearest MRT (X3) ranged from 23.38 to 6488.02, while the number of convenience stores (X4) ranged from 0 to 10. According to the two mentioned variables, there were significant differences in the data range of the mentioned

attributes. It implied that X3 was overly appreciated if being compared to X4, while X4 was considered to be neglected when being processed with X3. Consequently, there was an imbalance in the dataset where the data highly varied. Because of that reason, it was required for the dataset to be normalized using z-score normalization to convert the dataset into zero mean and unit variance distribution.

On the other hand, the normalized dataset showed that the data of each feature did not vary widely. Also, the average value of the normalized dataset was close to zero. In contrast, the data in the original dataset varied widely, and the mean of each attribute was in a great number. The plot of the original data of X3 and X4 is depicted in Figure 2. Whereas, the variable of X3 and X4 after being normalized is plotted in Figure 3.

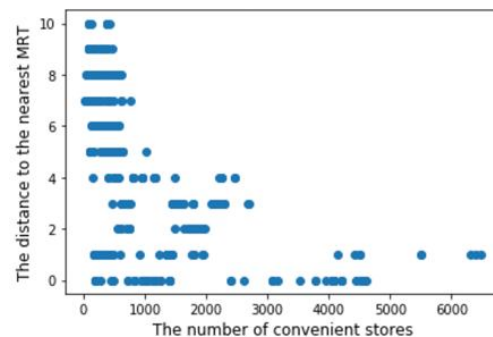


Figure 2: The plot of the number of convenient stores vs. the distance to the nearest MRT in the original dataset

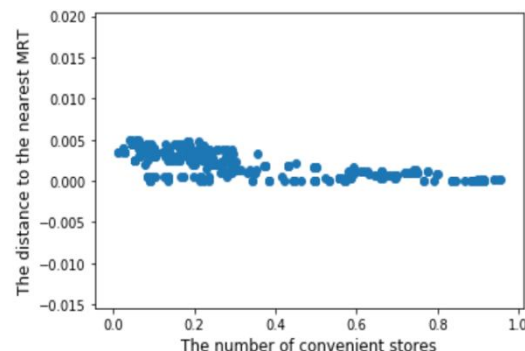


Figure 3: The plot of the number of convenient stores vs. the distance to the nearest MRT in the normalized dataset

After obtaining two datasets, e.g., original dataset and normalized dataset, the data splitting process was conducted for evaluation. Each of those datasets was separated into 75% of data training (310 data points) and 25% of data testing (104 data points). In the data testing, the true target was removed while the input data were inserted into the data model to obtain the estimated target. The data were split using a simple random sampling method.

After the data preprocessing phase was applied, the datasets were modeled in the data processing stage. Each of the datasets was processed using two different regression algorithms, e.g., linear regression and polynomial regression. The former was implemented in the training dataset for the original dataset and normalized dataset. The expected model of linear regression was the feature with optimum weight

added with intercept. In obtaining the optimum model, the model fitting was necessary to be employed. The term model fitting was defined as a process of model generalization in the learning process on a training dataset. The fitting benefited from minimizing the overfitting problem that might occur in the model. The overfitting might happen when it modeled the noise and outliers instead of the real data. The model was expected to be lack accuracy if the model underwent an overfitting problem [17].

After the regression line obtained from the original dataset, the regression weights for the original dataset were earned – shown in Table 2.

Table 2: Weights of Linear Regression Model for the original dataset

Variable	Weight
X1	5.149e+00
X2	-2.697e-01
X3	-4.488e-03
X4	1.133e+00
X5	2.255e+02
X6	-1.243e+01

According to the result, the value of the intercept was huge – approximately -14,441.99. In addition, from Table 2, it is shown that X1, X4, and X5 were correlated positively with the target value (y). It meant that when the variable X1, X4, or X5 increase, there was also an improvement in the target value (y). Conversely, X2, X3, and X6 had a negative correlation with the target. In other words, when the value of X2, X3, and X6 decreased, there was a rise in target value (y). Additionally, the linear regression algorithm was also applied to the normalized data. The model was then fitted for the normalized data – resulting in regression weight shown in Table 3.

Table 3: Weights of Linear Regression Model for the normalized dataset

Variable	Weight
X1	5.957e+00
X2	-2.574e-01
X3	-2.635e-02
X4	2.159e-01
X5	2.936e+02
X6	-1.590e+02

According to the result, the intercept of the linear regression model in the normalized dataset had a tiny value – about 0.0301, which had a huge difference with the intercept of the model of the original dataset. It was said that normalization was able to minimize the dispersion of the data. Additionally, the value of X1, X4, and X5 had a positive correlation with the target while (y) while the value of X2, X3, and X6 negatively correlated with the target value (y).

After the optimum model was earned, the model was required to be evaluated. The evaluation techniques employed were MSE and R-squared. Based on the result, the linear model,

when implemented in the original dataset, generated an MSE score of 103.45. It was shown that the MSE score was relatively high. Therefore, it was said that the data were widely dispersed from the central point (mean). Additionally, the other evaluation technique, i.e., R-squared, produced a score of 0.539. The R-squared score earned was fair, yet still not considered to be a high score.

Besides, in the normalized dataset, MSE and R-squared produced were 2.194×10^{-5} and 0.647, respectively. The value of MSE of the linear model when being applied in the normalized dataset was lower compared when being implemented in the original one. However, the value of the R-squared of the linear model for the normalized dataset was higher than the original dataset. It was concluded that in the linear regression model, z-score normalization was able to boost the accuracy score, specifically in MSE and R-squared evaluation.

The second algorithm to predict the data was the polynomial regression model. This technique contained every possible feature combination of the dataset. In this technique, the optimum degree was required to be achieved before implementing a polynomial regression model. In achieving the optimum order, MSE and R-squared scores were measured by executing every different polynomial order ranging from zero-order to 20-order. After that, the MSE and R-squared scored over different polynomial orders implemented in the original dataset are shown in Figure 4 and Figure 5, respectively.

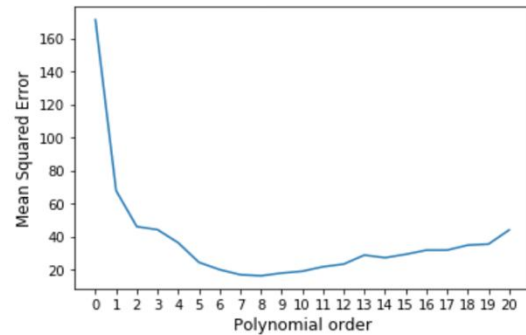


Figure 4: MSE of polynomial orders in the original dataset

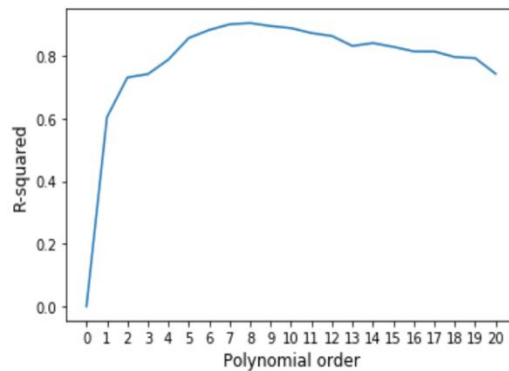


Figure 5: The R-squared score of polynomial orders in the original dataset

According to the result, it was observed that degree 2 produced an adequately low MSE score and relatively high R-squared. In more detail, this model MSE scored approximately 45.957, and R-squared scored about 0.731. Since the greater degrees produced better scores, it did not imply that higher degrees were able to generate a better model. The reason was that a model with higher degrees was prone to overfitting. Also, it might even model the noise or outlier rather than the real data. Therefore, it was summarized that the quadratic polynomial regression was attributed as the optimum model for the prediction. With the order = 2 and 6 features, the model contained 28 feature combinations of the input features, in addition to the intercept.

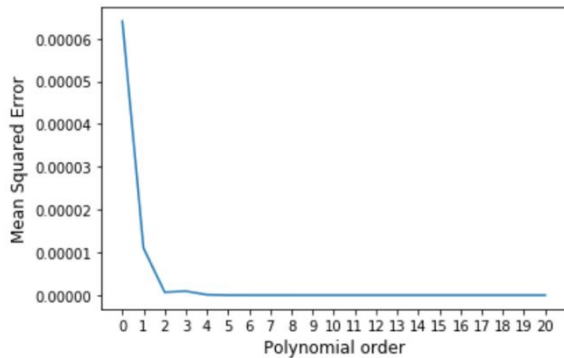


Figure 6: MSE of polynomial orders in the normalized dataset

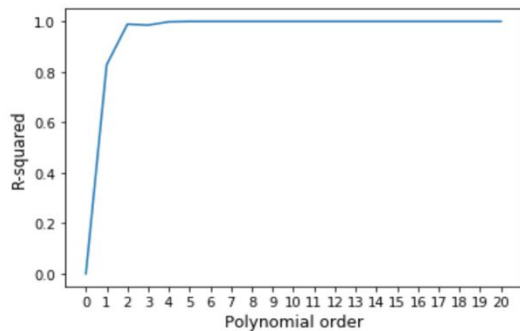


Figure 7: R-squared score of polynomial orders in normalized dataset

In addition to the original dataset, the quadratic polynomial regression was also chosen as the most suitable model for prediction in the normalized dataset. As observed in Figure 6, the MSE score reached the lowest point at order 2 (7.044 X 10⁻⁷). It, then, did not experience a noticeable decrease for the higher degrees. Furthermore, it was supported with a high R-squared score in the normalized dataset, which was around 0.989, as shown in Figure 7.

After the optimum order earned, the quadratic polynomial regression model for the original dataset and the normalized dataset were implemented. The weights of the quadratic polynomial regression model applied in the original dataset and the normalized dataset are informed in Table 4 and Table 5, respectively.

Table 4: Weights of Polynomial Regression Models for the original dataset

Variable	Weight
1	3.17819989e+02
X1	-4.10833004e+04
X2	4.01242662e+02
X3	1.86431697e+01
X4	1.01149970e+04
X5	2.39182596e+05
X6	1.16425500e+06
X1 ²	1.00419397e+01
X1 · X2	1.76928972e-01
X1 · X3	8.76217764e-04
X1 · X4	2.89803614e-01
X1 · X5	-3.59516273e+01
X1 · X6	1.27522736e+01
X2 ²	2.10246050e-02
X2 · X3	-1.72036670e-05
X2 · X4	2.45231003e-03
X2 · X5	-7.27596176e+00
X2 · X6	-4.74596006e+00
X3 ²	-1.04957534e-06
X3 · X4	-1.79868851e-03
X3 · X5	-2.90231521e-01
X3 · X6	-1.08278182e-01
X4 ²	6.61144505e-03
X4 · X5	-1.27322477e+02
X4 · X6	-6.18500729e+01
X5 ²	4.77790067e+03
X5 · X6	-3.32588307e+03
X6 ²	-4.55218254e+03

Table 5: Weights of Polynomial Regression Models for the normalized dataset

Variable	Weight
1	-4.57650190e-01
X1	9.38778893e+00
X2	3.40327784e-01
X3	3.48669934e-02
X4	-8.75760114e-01
X5	2.25211728e+01
X6	-1.59047520e+02
X1 ²	2.69540836e+03
X1 · X2	2.35351049e+01
X1 · X3	-3.75950277e+00
X1 · X4	-5.56707153e+02
X1 · X5	1.88163131e+05
X1 · X6	-1.28781556e+05
X2 ²	-2.25194978e+01
X2 · X3	-3.55023732e-01
X2 · X4	-8.12054171e+00
X2 · X5	8.20857762e+03
X2 · X6	-2.08062783e+03
X3 ²	-2.10973748e+01
X3 · X4	2.11541630e-01
X3 · X5	-2.95979741e+01

X3 · X6	6.78663123e+01
X4 ²	-2.13204511e+01
X4 · X5	-4.93819004e+04
X4 · X6	1.93829131e+04
X5 ²	2.28680490e+06
X5 · X6	-4.05454456e+06
X6 ²	1.48394295e+06

In addition to the weights, the intercept of the quadratic polynomial regression accounted for -27391659.15, while the intercept of the normalized dataset was 21.51. Furthermore, the weight of each variable is shown in Table 4 and Table 5 for the original dataset and normalized dataset, respectively. According to the weight data, it was observed that normalization was able to noticeably change the pattern of the regression model in terms of weights. In other words, there were several features weighted positively concerning the target value. However, after implementing z-score normalization, the weight scores became negative. In some points, it also applied to the negative weights that were converted into positive. Although some were changed, there were still some other weights that remained unchanged. This issue did not only apply to the quadratic polynomial regression but also in the linear regression model.

Finally, the quadratic polynomial regression model was evaluated using MSE and R-squared. According to the result, the quadratic linear regression model for the original dataset generated an MSE score of 45.95, which was high. On top of that, R-squared also produced a score of 0.731. From the R-squared score generated, it implied that the model was considered as an acceptable model, but the value still required improvement. In the normalized dataset, this technique performed better according to the evaluation results. It resulted in an MSE score of 7.044×10^{-7} and an R-squared score of 0.989. It was concluded that the quadratic polynomial regression performed better in the normalized dataset than in the original dataset.

4. CONCLUSION

Based on the result, the quadratic polynomial regression outperformed the linear regression to predict the real estate dataset in Sindian District, New Taipei City, Taiwan. In the normalized dataset, the former earned low scores of MSE and R-squared score, which were 7.044×10^{-7} and 0.989, respectively. Whereas, it also produced 45.957 for MSE and 0.731 R-squared for the original dataset. The linear regression – in the normalized dataset – earned 2.194×10^{-5} for the MSE score and 0.647 for the R-squared score. Besides, in the original dataset, it generated 103.45 for the MSE score and 0.539 for the R-squared score.

Based on the result, it was summarized that z-score normalization could successfully improve the accuracy score in the aspects of MSE score and R-squared score. In other words, the data were required to be rescaled into new distribution, i.e., zero-mean and unit-variance distribution, to obtain better accuracy. Thus, it was concluded that the quadratic polynomial regression possessed the best fit of the

model for predicting the real estate price. Moreover, if the algorithm combined with the z-score normalization, the accuracy score produced becomes even higher.

Nevertheless, in this research, the polynomial regression method utilized many feature combinations. In total, there were 28 feature combinations from the original six features. According to Table 4 and Table 5, not all features had sufficient weight for predicting the target. It might also consume a high amount of resources and time to process the large dataset. As a proposed solution, in the future works, it is necessary to apply a regularization method that benefits from reducing features that has a small amount of weight. Hopefully, the future works possess a less complex computation compared to the current research.

REFERENCES

- [1] Yu, C.M. and Chen, P.F., 2018. **House Prices, Mortgage Rate, and Policy: Megadata Analysis in Taipei.** *Sustainability*, 10(4), p.926. <https://doi.org/10.3390/su10040926>
- [2] Erraisi, A., Belangour, A., 2019. **Meta-modeling of Big Data management layer.** *International Journal of Emerging Trends in Engineering Research*. vol. 7, no. 7, pp. 36-43. <https://doi.org/10.30534/ijeter/2019/01772019>
- [3] Kumar, N.V.S.P., Sastry, J.K.R., Rao, K.R.S., 2019. **Mining Negative Associations between Regular and Frequent Patterns hidden in Static Databases.** *International Journal of Emerging Trends in Engineering Research*. vol. 7, no. 8, pp. 54-67. <https://doi.org/10.30534/ijeter/2019/04772019>
- [4] Ge, Y. and Wu, H., 2019. **Prediction of corn price fluctuation based on multiple linear regression analysis model under big data.** *Neural Computing and Applications*, pp.1-13.
- [5] Wei, J., Chen, T., Liu, G. and Yang, J., 2016. **Higher-order multivariable polynomial regression to estimate human affective states.** *Scientific reports*, 6, p.23384. <https://doi.org/10.1038/srep23384>
- [6] Botchkarev, A., 2018. **Evaluating performance of regression machine learning models using multiple error metrics in Azure Machine Learning Studio.** *Available at SSRN 3177507*.
- [7] Prairie, Y.T., 1996. **Evaluating the predictive power of regression models.** *Canadian Journal of Fisheries and Aquatic Sciences*, 53(3), pp.490-492. <https://doi.org/10.1139/f95-204>
- [8] Dinç, İ., Sigdel, M., Dinç, S., Sigdel, M.S., Pusey, M.L. and Aygün, R.S., 2014, March. **Evaluation of normalization and pca on the performance of classifiers for protein crystallization images.** In *IEEE SOUTHEASTCON 2014* (pp. 1-6). IEEE.
- [9] Jeon, J., 2015. **The strengths and limitations of the statistical modeling of complex social phenomenon: Focusing on SEM, path analysis, or multiple regression models.** *Int J Soc Behav Educ Econ Bus Ind Eng*, 9(5), pp.1594-1602.

- [10] Lee, K.Y., Kim, K.H., Kang, J.J., Choi, S.J., Im, Y.S., Lee, Y.D. and Lim, Y.S., 2017. **Comparison and analysis of linear regression and artificial neural network.** *International Journal of Applied Engineering Research*, 12(20), pp.9820-9825.
- [11] Harrington, P., 2012. *Machine learning in action.* Manning Publications Co..
- [12] Royston, P., Ambler, G. and Sauerbrei, W., 1999. **The use of fractional polynomials to model continuous risk variables in epidemiology.** *International journal of epidemiology*, 28(5), pp.964-974.
<https://doi.org/10.1093/ije/28.5.964>
- [13] Emmert-Streib, F. and Dehmer, M., 2019. **Evaluation of Regression Models: Model Assessment, Model Selection and Generalization Error.** *Machine Learning and Knowledge Extraction*, 1(1), pp.521-551.
<https://doi.org/10.3390/make1010032>
- [14] Ekpenyong, E.J., Okonnah, M.I. and John, E.D., 2008. **Polynomial (non linear) regression method for improved estimation based on sampling.** *J. Applied Sci*, 8, pp.1597-1599.
<https://doi.org/10.3923/jas.2008.1597.1599>
- [15] Hofwing, M., Strömberg, N. and Tapankov, M., 2011. **Optimal polynomial regression models by using a genetic algorithm.** In *The Second International Conference on Soft Computing Technology in Civil, Structural and Environmental Engineering Conference, 6-9 September, Chania, Crete, Greece.*
- [16] Dey, A., 2016. **Machine learning algorithms: a review.** *International Journal of Computer Science and Information Technologies*, 7(3), pp.1174-1179.
- [17] Cai, Y., Zheng, W., Zhang, X., Zhangzhong, L., & Xue, X. (2019). **Research on soil moisture prediction model based on deep learning.** *PloS one*, 14(4), e0214508.
<https://doi.org/10.1371/journal.pone.0214508>