Effect of the Multicollinearity of Interaction Terms on the Performance of the ANN Model

Celso FERNANDO [a], Toshio YOSHII [b], Takahiro TSUBOTA [c], Hirotoshi SHIRAYANAGI [d]

[a,b,c,d] Graduate School of Science and Engineering, Ehime University, Ehime, 790-8577, Japan
[a] E-mail: celso.fernando.12@cee.ehime-u.ac.jp
[b] E-mail: yoshii@cee.ehime-u.ac.jp
[c] E-mail: t.tsubota@cee.ehime-u.ac.jp
[d] E-mail: shirayanagi@cee.ehime-u.ac.jp

Abstract: This research analyzes the effect of adding an interaction term to the ANN model. The analysis focuses on how the multicollinearity between the interaction term and other input variables affect the performance of the ANN model. The performance of the model is measured based on the ROC-AUC and the multicollinearity is measured based on the variance inflation factor, VIF, of the interaction term. The result shows that the interaction term with high VIF worsens ANN performance.

Keywords: ANN, ROC-AUC, Interaction terms, Multicollinearity

1. INTRODUCTION

During the past decades, the collection and storage capability of the data has largely increased in all fields of science. This phenomenon challenges the researchers working in different domains including engineering, mainly because the traditional statistical methods of analysis fail to deal with a large volume of data (Azhagusundari and Thanamani, 2013). This difficulty is increased when the dataset presents a very complex relationship pattern between the explanatory and the response variables, in other words, when the dataset can not fit in any statistical assumption governing a data distribution (Gardner and Dorling, 1998; Lin et al., 2015). It is at this domain where the artificial neural network, ANN, outperforms the traditional statistical method in the analysis of the different fields of knowledge (Li and Cao, 2018; Snieder et al., 2019).

The ANN consists of interconnection between neurons from different layers, and these are responsible for mapping non-linear relationships between the input vector and an output. The ANNs are known for their high prediction performance even when there is a very complex relationship.

Even though the ANN can deal with a large volume of data, several studies have focused on the selection of the most important variables for ANN prediction analysis (Yao et al., 1998; Gevrey et al., 2003; Nourani et al., 2012). It is because, in such large datasets, not all the available variables are important for the analysis, and they may represent a computational cost.

However, even if all variables may have some importance, it is still of interest in many applications to minimize the number of variables in any modelling. To reduce the number of variables in data modelling without harm the performance of the model has many benefits, for example, (1) if the dataset has to be purchased, the researcher will want to obtain a minimum amount of data which can maximize the performance of the analysis; (2) it reduces the computational cost, Ben Meskina (2013); (3) allows the model to find more efficient and

useful patterns (Azhagusundari and Thanamani, 2013; Ben Meskina, 2013) and; (4) it reduces the complexity of the model and easy the interpretability, RJ and AM (1996).

However, this research analyzes the effect of the addition of an interaction term in the performance of the ANN model. An interaction occurs when the relation between the existing variables, explanatory and the response variables, changes as a function of other(s) variable(s), Fürst and Ghisletta (2009).

There are fewer studies on the interaction term in the ANN. Since ANN can model non-linear relationship as well as can find complex patterns among the variables, De Veaux and Ungar (1994), the existing studies focus on detecting whether there is or not interaction effect between the input variables through the analysis of the connection weights between the nodes from different layers (Tsang et al., 2018; Liu and Wang, 2020). While the algorithm proposed by Tsang et al. (2018) and Liu and Wang (2020), the Neural Interaction Detection (NID) and the Persistence Interaction Detection (PID), respectively, aim to discover the combined effects among the input variables in the output of the ANN, the goal of this paper is on the model performance when the interaction term is added to the input variable. In other words, this paper investigates whether the addition of an interaction term to the input variables of the ANN model affects the model performance. This is very different from the previous studies by Tsang et al. (2018) and Liu and Wang (2020).

This study focuses mainly on model performance as a consequence of the level of multicollinearity between the interaction term and the other explanatory variables.

The multicollinearity refers to non-independence between the input variables and, describes the situation where two or more explanatory variables are statistically linearly related or they have a homogeneous characteristic. Its analysis is important due to the reproducibility of the statistical evidence of a relationship between the variables and the output, mainly when a model is trained on data from a place and time, and used to predict data from different place and time (Dormann et al., 2013; Lindner et al., 2020). The multicollinearity is a sample characteristic problem.

P. Obite et al., (2020) compared how the ANN model and OLSR models perform in the presence of multicollinearity; His finding was that the ANN model performs better than the OLSR model. However, he did not analyze how two ANN models perform when one is exposed to multicollinearity and another is not, or when the two ANN models are exposed to different levels of multicollinearity.

Therefore, this study explores this gap by investigating the effect of the multicollinearity of interaction term on the performance of the ANN model. The performance of the ANN model is evaluated based on the receiver operator characteristics area under the curve, ROC-AUC, while the multicollinearity is measured by the variance inflation factor, VIF, of the interaction term. The expected result is that the higher the VIF of the interaction term, the less forecasting ability of the model where it is added.

Empirical traffic accident data are used in the research. The interaction terms, among the available variables, are generated based on the association rule mining approach. In this research, the interaction terms are defined as the maximal itemset.

Each maximal itemset, which corresponds to the interaction term, is added to the standard model one at a time, and the performance of the models with data containing each interaction term are compared.

This paper is organized as follow, in the first section a brief introduction of the research. The second focus on the methodology, here, all the procedure of data pre-processing for the analysis are explained. The third section presents the case study. Empirical data are used for the verification of the analysis. The fourth and fifth sections are the results and discussion, respectively. The analysis finishes with the conclusion in the sixth section.

2. DEFINITION OF INTERACTION TERM

This section details the steps for interaction terms definition, the ANN models performance and variance inflation factor estimation.

In the real-life, most phenomena (e.g. road traffic accident, cancer disease, etc) are not only dependent on one factor but the combination of factors. It is, therefore, important to understand the effect of the combination of factors (interaction effect) in the output.

However, when the available variables are continuous, the interaction term definition is a challenging task. One way to overcome such a problem is the categorization of the continuous variables.

2.1 Expectation-Maximization EM-Algorithm for Categorization of the Continuous Variables

In this paper, the EM algorithm is applied to cluster every single continuous variable. The EM algorithm iteratively adjusts the initial mixture model parameter, assigning probabilities of cluster membership for each data until reaching the local optimal solution which better fit the data distribution. The cluster membership, in other words, the category of the data is defined based on the highest probability assigned to the data. (Bradley, Fayyad and Reina, 1998)

The categorization of the data allows converting the initial data (standard data) to categorized data, which will be used in the following steps.

Let the Standard data consists of X_{ij}, $i=1,\ldots,p$ explanatory variables where all or part of them are continuous, and $j=1, \ldots, n$ data instances. Then, the categorized data consists of X^{cat}_{ij}, where all explanatory variables are categorical.

2.2 Association Rule Mining

Association rule is a conditional rule between attributes occurrence in the dataset. They are represented in the form of A→B, translated as "a significant proportion of data instances contain the item A together with the item B". Item A represents a set or subset of the explanatory variables from the categorized data, X^{cat}_{ij}; they consist of at least 2 explanatory variables. Item B is the target response variable. An association, A→B, is considered as a rule if its occurrence in the database is at least equal to a given threshold.

Association rules have been used to extract the sets or subsets of data that most often occur together, so-called frequent patterns or rules. In this study, the features from the explanatory variables that most often occur with the target response are the frequent patterns. The frequent patterns are, therefore, considered as the candidate for interaction terms. Some problems from the association rule approach at identifying the frequent patterns are: (1) the definition of the appropriate threshold, since setting too high a value may lead to the miss of important relations, on the other hand, setting too low a value may lead to extract irrelevant rules. (2) the difficulty in the interpretation and management of a huge number of rules. (3) the redundant rules, those rules in which their information are contained in other rules. These problems have been mentioned in studies like (Brijs, Vanhoof and Wets, 2003; Pasquier, 2009; Rajalakshmi et al., 2011). However, some of these problems can be solved by extracting only the maximal itemsets.

In this paper, the maximal itemsets are considered as interaction terms. Let the interaction terms be represented as At. For a better understanding of the maximal itemset concept, we refer the reader to Rajalakshmi et al., 2011, and page 47-48 of Aggarwal and Han, 2014.

After the extraction of the interaction terms, the ANN models are developed.

2.3 ANN Model and Performance Metrics

The artificial neural network, ANN, is used for modelling. The data used for modelling is the standard data, Xi - continuous variable. One interaction term, At, is added to the input variable one at a time. Thus, the models are distinguished from one another by the interaction term which is added to the input variable, and they are represented as *standard+At*. The standard model is the model which does not have any interaction term in the input and, it is simply represented as Standard.

The model's performance is measured by the receiver operator characteristics area under the curve, ROC-AUC. The ROC-AUC is a unitless index, and it represents the probability that randomly selected data instances that experienced the target response will have a higher predicted probability of having the target response as the outcome compared to a randomly selected data instance that did not experience the target event (Austin and Steyerberg, 2012).

Each model is run several times (an equal number of time for all models) and the maximum ROC-AUC is collected for comparison. Let ROC-AUC$_{Std}$ and ROC-AUC$_{t}$ represent the maximum ROC-AUC observed from the running of the standard model and *standard+At* model, respectively.

The expected result is that the model which contain the interaction term, At, with low VIF will perform better than those with high VIF, Murray et al., (2012).

2.4 Multicollinearity among the Input Variables - VIF

The multicollinearity is measured by the variance inflation factor, VIF, which estimates the amount of inflation on the error terms in a linear regression model (Lindner, Puck and Verbeke, 2020). Each explanatory variable, at a time, is assumed to be the response variable of multiple linear regression where other explanatory variables remain as explanatory variables. The fitness of the multiple linear regression, R^2, obtained by regressing each variable on the remaining variables is used to calculate the VIF, it is, for a multiple regression model with p predictors, X_i, $i=1, \ldots, p$, VIF of the *ith* predictor variable can be expressed by

$$VIF_i = \frac{1}{1-R^2_i}, i = 1, \ldots, p \tag{1}$$

Where R^2_i corresponds the fitness of the multiple linear regression when the *ith* predictor is the response variable and all other inputs are the explanatory variables, Murray et al. (2012).

This study focuses on the effect of the multicollinearity between the interaction term and the other input variables in the performance of the ANN model and the VIF of the interaction term is what matters for the analysis. Thus, for each model containing an interaction term *At* in its input, the VIF$_t$, is estimated by Equation 1. Where, $t = i$ and $t=1, \ldots p$.

Several researchers have reported some effect of multicollinearity in the nonlinear multivariate analysis (Wonsuk et al., 2013; Vu, Muttaqi and Agalgaonkar, 2015). Murray et al. (2012), argue that models with high collinearity have estimators with lower precision,

consequently, with problems in testing hypotheses and forecasting.

2.5 Comparison Analysis

The effect of the multicollinearity of the interaction terms on the performance of the ANN model is observed by plotting the ROC-AUC of the models and the VIF of the interaction terms added to the model's input.

3. CASE STUDY

The study uses empirical road traffic accident data from the Hanshin expressway in Osaka, Japan.

3.1 Data Description

The April 2010 – March 2016 data used in the analysis were collected in three routes (Higashi-Osaka, Ikeda and Matsubara) from the Hanshin expressway in Osaka, Japan.

The road traffic data consist of variables such as traffic volume, time-mean speed and occupancy which were collected at each 5-minutes time resolution and the spatial resolution corresponds to the distance between two counter-devices. The road geometry data (i.e. horizontal and vertical alignment) of the road segment were collected in a spatial resolution of 100-meters; The precipitation is reported in the hourly average.

Since the road traffic data and the precipitation data were collected at a different time resolution, an aggregation process to uniformize the time resolution has been done. Therefore, all road traffic data were aggregated for a time resolution of one hour as shown in Equations 2 to 4.

Traffic volume: Two traffic volume exist in the data, one refers to the general traffic volume (*i.e.,* all vehicle types in the dataset, this is denoted as 'Trf'), and the other refers to a specific traffic volume, where only heavy vehicles are counted (denoted as 'HVT'). The aggregation of both traffic data is based on Equation 2.

$$Q_{s,h} = \sum_{l}^{N^s} \sum_{n}^{12} q_{s,h,l,n}^{5\min} \tag{2}$$

where,

N^s : the total number of lanes.

n : each 5-minute interval within one-hour h. (each one hour consists of 12n)

s : road section (covered by each loop counter device).

$q_{s,h,l,n}^{5\min}$: the traffic volume at road segment s in the lane l at each 5-minute n in each time h.

$Q_{s,h}$: the hourly traffic volume in the section s at time h. it can be $Q_{s,h}^{Trf}$ and $Q_{s,h}^{HVT}$ refereeing to general traffic and heavy-vehicle traffic, respectively.

Time-mean Speed: The speed was calculated as an instantaneous speed observed at each location (*i.e.,* at loop counter device location) as defined in the following equation

$$v_{s,h}^{hour} = \frac{\sum_{l}^{N^s}\sum_{n}^{12} v_{s,h,l,n}^{5\min}}{12N^s} \tag{3}$$

where,

$v_{s,h}^{hour}$: the average time-mean speed in the lane l from the road section s, at 5-minutes of a given

$v_{s,h,l,n}^{5\min}$: the average hourly time-mean speed in the road section s of a given hour h.

Occupancy: The occupancy of each segment within each five-minute interval was averaged into an hourly time interval based on the following equation

$$occup_{s,h}^{hour} = \frac{\sum_{l}^{N^s}\sum_{n}^{12} occup_{s,h,l,n}^{5\min}}{12N^s} \tag{4}$$

where,

$occup_{s,h}^{hour}$: the hourly occupancy in the section s at time h.

$occup_{s,h,l,n}^{5\min}$: the 5-minutes occupancy in lane l from the section s, at a given time h.

3.2 Database Construction

Since the road traffic data were collected from each lane l and 5-minute interval and aggregated for each road section s and one-hour interval h. We, also, calculated the standard deviation of each explanatory variable, within the hour interval and road section. Therefore, after the aggregation process, the available road traffic data consists of average hourly traffic volume (whole traffic and heavy-vehicle traffic), the standard deviation of the hourly traffic volume (whole traffic and heavy-vehicle traffic), hourly time-mean speed, the standard deviation of the hourly time-mean speed, hourly occupancy, and the penetration rate of the heavy-vehicle in the traffic volume.

In total, 11 explanatory variables are available, where 8 are continuous variables and 3 are categorical variables.

The response variable is binary data, which describes whether or not an accident has been observed, taking 1 if at least an accident has occurred, 0 otherwise.

To allow the implementation of data mining for maximal itemset extraction, the EM algorithm was applied to group each continuous variable into categories. The result is shown in Table 1.

After the categorization of the explanatory variables, the maximal itemset (hereafter, interaction term) are extracted applying the apriori algorithm, a package from the R-studio programme. As explained before, the interaction terms are the set or subset of the explanatory variables which most often occur with the response variable, in our case *road traffic accident*.

Nine interaction terms were extracted from the dataset. Table 2 shows the result of data mining and its descriptions.

Table 1. Data categorization

Description (units)	Categories [interval of the data]
Traffic volume (Veh/hour)	**a1** (0, 31], **a2** [32, 160], **a3** [161, 535], **a4** [536, 2727]
Heavy-vehicle traffic volume (Veh/hour)	**b1** (0, 4], **b2** [5, 27], **b3** [28, 97], **b4** [98, 1163]
Traffic volume deviation (Veh)	**d1** (0, 4], **d2** [4.1, 19.4], **d3** [19.5, 102.1]
Heavy-vehicle traffic deviation (Veh)	**e1** (0, 1.7], **e2** [1.8, 8.4], **e3** [8.5, 43.2]
Time-mean speed (Km/hour)	**g1** (0, 63.5], **g2** [63.6, 90.8], **g3** [90.9, 122]
Time-mean speed deviation (Km/hour)	**h1** (0, 8.2], **h2** [8.3, 22.9], **h3** [23, 63.2]
Occupancy (%)	**j1** (0, 2], **j2** [2.1, 96.2]
Heavy-vehicle Rate (%)	**k1** (0, 0.088], **k2** [0.089, 0.28], **k3** [0.29, 0.43]
Precipitation	**p1** (Sunny), **p2** (Rain)
Vertical road alignment (Slope)	**q1** (Upward), **q2** (Flat), **q3** (Down)
Horizontal geometry	**l1** (Straight), **l2** (Curve), **l3** (Tight)

Table 2. Interaction terms (maximal itemsets)

Interaction terms (A_t)	Description (Driving under the following condition)
k2p1	A normal rate of the heavy vehicle under the sunny condition
l1p1	The straight segment under the sunny condition
d2p1	Low traffic volume deviation within one-hour under the sunny condition
e2j2p1	Normal traffic volume deviation of heavy-vehicle within one-hour, high occupancy, and sunny condition
a4j2p1	High traffic volume, high occupancy, and sunny condition
e2j2k2	Normal traffic volume deviation of heavy-vehicle within one-hour, high occupancy, and normal heavy-vehicle penetration rate
a4j2e2	High traffic volume, high occupancy, and normal traffic volume deviation of heavy-vehicle within one-hour
a4j2k2	High traffic volume, high occupancy, and normal heavy-vehicle penetration rate
b4j2	High heavy-vehicle traffic and high occupancy

3.3 ANN Modeling

Ten models were developed. The first model is the baseline. This consist only of the standard data; no interaction term is added to the input. Therefore, this model takes the name of the standard model.

The other nine models consist of the addition of one interaction term to the standard data, in other words, they consist of the standard explanatory variables plus one interaction term at a time. These models take the name of the interaction term which is added to the model input variable (*e.g., standard+At*).

While the standard model consists of eleven explanatory variables, the other nine models consist of twelve explanatory variables.

The models' output the probability of each data instance to be classified as an accident-prone event or not. The models are trained with 80% of the data and test on 20%, the objective function is binary cross-entropy.

For each data, 24 different ANNs where executed. The architecture of the ANNs differs from the number of neurons in the hidden layers and the batch-size. All the ANNs have two hidden layers, where the number of neurons varies as follow: (8, 10, 11, 12) in the first hidden layer, and (8, 6, 4) in the second hidden layer. Two training are executed for each ANN architecture, each of them taking a different batch-size (256, 512). The ROC-AUC of each

ANNs is recorded and compared. For each data (*i.e.,* the standard data and other nine data), the highest value of ROC-AUC is selected for analysis.

On the other hand, for each *standard+At* data, the analysis of multicollinearity between the interaction term, *At*, with other explanatory variables, X_i, is performed.

Then the performance of the models (*standard+At*) and the level of the multicollinearity of the interaction terms (VIF of the *At*) are plotted.

4. RESULTS

Figure 1 shows the relationship between the VIF of the interaction terms and the performance of the ANN model when the interaction terms are used as one of the explanatory variables (i.e., *standard+At* models).

Since the *standard model* does not have an interaction term in the input, it is represented in Figure 1 as a horizontal dash-line. The line indicates the ROC-AUC score.

The results show the higher the VIF of the interaction term the lower the performance of the model where the interaction term is part of the input variable.

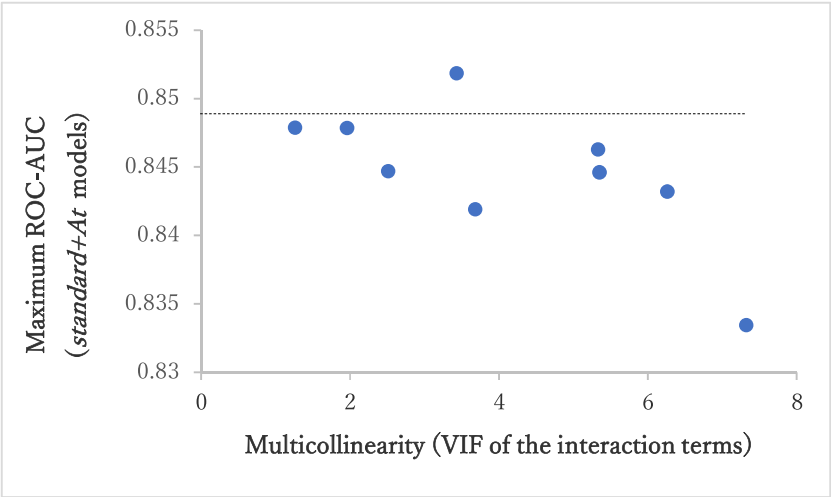

Figure 1. Relationship between VIF and ROC-AUC

5. DISCUSSION

The higher the ROC-AUC of the model with the interaction term the lower the VIF of the interaction term. These results agree with the explanation of Austin and Steyerberg, 2012; Murray et al., 2012; Lindner et al., 2020.

According to Austin and Steyerberg (2012), the ability of a model to distinguish two potential groups (model performance in classification task) is frequently assessed using the ROC-AUC.

In a binary univariate model, Austin and Steyerberg (2012) applied an analytical analysis to demonstrate that ROC-AUC is a function of the means and the variances of the explanatory variable in both groups of the outcome. Austin and Steyerberg argued that the ROC-AUC increases as the difference in the mean of the explanatory variable between both groups increases. However, for a multivariate model, the high predictive ability is expected in

a model containing independent explanatory variables that are strongly associated with the outcome, in other words, the high predictive ability is expected in more heterogeneous samples.

Lindner et al., (2020), states that VIF represents the degree to which the standard error of the predicted response is too large, it is, the average distance from the predicted regression line given by the variable in interest to the observed points is relatively larger than that if the collinearity does not exist or is small. This statement is also supported by Murray et al. (2012), who argue that higher VIF reflects an increase in the variance of the estimated regression coefficient, therefore, models with high collinearity, high VIF, have estimators with lower precision, consequent low forecasting ability.

6. CONCLUSION

This research analyzes the effect of adding an interaction term to the ANN input and the model's performance. The analysis focuses on how the multicollinearity between the interaction term and other input variables affect the performance of the ANN model. The performance of the model is measured by the ROC-AUC and the VIF describes the multicollinearity level between the interaction term and other explanatory variables.

From the result of the experiment, we conclude:

(1) interaction terms with a high VIF value worsen the ANN performance compared to the interaction terms with low VIF values.

(2) the addition of an interaction term to the ANN does not necessarily imply the improvement of the ANN model performance compared to the standard model.

(3) the categorization of the input variables for interaction term definition may affect the performance of the ANN even if the VIF is low.

ACKNOWLEDGEMENTS

The authors would like to thank Hanshin expressway Co., Ltd for providing the data for this research.

REFERENCE

Aggarwal, C. C. and Han, J. (2014) Frequent Pattern Mining. New York: Springer International Publishing. DOI: 10.1007/978-3-319-07821-2.

Austin, P. C. and Steyerberg, E. W. (2012) 'Interpreting the concordance statistic of a logistic regression model: Relation to the variance and odds ratio of a continuous explanatory variable', BMC Medical Research Methodology. BMC Medical Research Methodology, 12(1), p. 1. DOI: 10.1186/1471-2288-12-82.

Azhagusundari, B. and Thanamani, A. S. (2013) 'Feature Selection based on Information Gain', International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2(2), pp. 18–21. DOI: 2278-3075.

Bradley, P. S., Fayyad, U. M. and Reina, C. (1998) 'Scaling {EM} Clustering to Large Databases', (MSR-TR-98-35).

Brijs, T., Vanhoof, K. and Wets, G. (2003) 'Defining Interestigness for Association Rules', BulDML at Institute of Mathematics and Informatics, 10(4), pp. 370–375.

Dormann, C. F. et al. (2013) 'Collinearity: A review of methods to deal with it and a simulation study evaluating their performance', Ecography, 36(1), pp. 27–46. DOI: 10.1111/j.1600-0587.2012.07348.x.

Fürst, G. and Ghisletta, P. (2009) 'Statistical Interaction between Two Continuous (Latent) Variables', 11th Congress of the Swiss Psychological Society August, pp. 1–12.

Gardner, M. W. and Dorling, S. R. (1998) 'Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences, Atmospheric Environment, 32(14–15), pp. 2627–2636. DOI: 10.1016/S1352-2310(97)00447-0.

Gevrey, M., Dimopoulos, I. and Lek, S. (2003) 'Review and comparison of methods to study the contribution of variables in artificial neural network models, Ecological Modelling, 160(3), pp. 249–264. DOI: 10.1016/S0304-3800(02)00257-0.

Li, Y. and Cao, H. (2018) 'Prediction for Tourism Flow based on LSTM Neural Network', Procedia Computer Science. Elsevier B.V., 129, pp. 277–283. DOI: 10.1016/j.procs.2018.03.076.

Lin, L., Wang, Q. and Sadek, A. W. (2015) 'A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction', Transportation Research Part C: Emerging Technologies. Elsevier Ltd, 55, pp. 444–459. DOI: 10.1016/j.trc.2015.03.015.

Lindner, T., Puck, J. and Verbeke, A. (2020) 'Misconceptions about multicollinearity in international business research: Identification, consequences, and remedies', Journal of International Business Studies. Palgrave Macmillan UK, 51(3), pp. 283–298. DOI: 10.1057/s41267-019-00257-1.

Liu, Z. and Wang, T. (2020) 'Towards Interaction Detection Using Topological Analysis on Neural Networks, (NeurIPS).

Ben Meskina, S. (2013) 'On the effect of data reduction on classification accuracy', 3rd International Conference on Information Technology and e-Services, ICITeS 2013. DOI: 10.1109/ICITeS.2013.6624071.

Murray, L. et al. (2012) 'VARIANCE INFLATION FACTORS IN REGRESSION MODELS WITH', Conference on Applied Statistics in Agriculture. DOI: 10.4148/2475-7772.1034.

Nourani, V. and Sayyah Fard, M. (2012) 'Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes', Advances in Engineering Software, 47(1), pp. 127–146. DOI: 10.1016/j.advengsoft.2011.12.014.

P. Obite, C. et al. (2020) 'Multicollinearity Effect in Regression Analysis: A Feed-Forward Artificial Neural Network Approach', Asian Journal of Probability and Statistics, 6(1), pp. 22–33. DOI: 10.9734/ajpas/2020/v6i130151.

Pasquier, N. (2009) 'Frequent closed itemsets based condensed representations for association rules', Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction, pp. 246–271. DOI: 10.4018/978-1-60566-404-0.ch013.

Rajalakshmi, M., Purusothaman, T. and Nedunchezhian, R. (2011) 'Maximal Frequent Itemset Generation Using Segmentation Approach, International Journal of Database Management Systems, 3(3), pp. 19–32. DOI: 10.5121/ijdms.2011.3302.

RJ, B. and AM, T. (1996) 'Choosing the best model: Level of detail, complexity, and model performance', Mathematical and Computer Modelling, 24(4), pp. 1–14. Available at: http://dx.doi.org/10.1016/0895-7177(96)00103-3.

Snieder, E., Shakir, R. and Khan, U. T. (2019) 'A Comprehensive Comparison of Four Input Variable Selection Methods for Artificial Neural Network Flow Forecasting Models', J Hydrol. Elsevier B.V., p. 124299. DOI: 10.1016/j.jhydrol.2019.124299.

Tsang, M., Cheng, D. and Liu, Y. (2018) 'Detecting Statistical Interactions from Neural Network weights', ICLR.

De Veaux, R. D. and Ungar, L. H. (1994) 'Multicollinearity: A tale of two nonparametric regressions', pp. 393–402. DOI: 10.1007/978-1-4612-2660-4_40.

Vu, D. H., Muttaqi, K. M. and Agalgaonkar, A. P. (2015) 'A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables', Applied Energy, 140, pp. 385–394. DOI: 10.1016/j.apenergy.2014.12.011.

Wonsuk, Y. et al. (2013) 'A Study of Effects of MultiCollinearity in the Multivariable Analysis', International journal of applied science and technology, 6(8), pp. 9–19. DOI: 10.1021/nn300902w.Release.

Woods, K. and Bowyer, K. W. (1997) 'Generating ROC curves for artificial neural networks, IEEE Transactions on Medical Imaging, 16(3), pp. 329–337. DOI: 10.1109/42.585767.

Yao, J. et al. (1998) 'Forecasting and analysis of marketing data using neural networks', Journal of Information Science and Engineering, 14(4), pp. 843–862. DOI: 10.6688/JISE.1998.14.4.7.