

PAPER • OPEN ACCESS

Feature Selection Using Multivariate Adaptive Regression Splines in Telecommunication Fraud Detection

To cite this article: Muhalim Mohamed Amin *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **864** 012059

View the [article online](#) for updates and enhancements.

You may also like

- [Parameter Estimation of Multivariate Adaptive Regression Spline \(MARS\) with Stepwise Approach to Multi Drug-Resistant Tuberculosis \(MDR-TB\) Modeling in Lamongan Regency](#)
S D P Yasmirullah, B W Otok, J D T Purnomo et al.
- [Modeling of toddler stunting in the province of east nusa tenggara using multivariate adaptive regression splines \(mars\) method](#)
D M Azizah and E O Permatasari
- [Multivariate Adaptive Generalized Poisson Regression Spline \(MAGPRS\) on the number of acute respiratory infection infants](#)
B W Otok, S Hidayati and Purhadi

Recent citations

- [A deep learning based traffic crash severity prediction framework](#)
Md Adilur Rahim and Hany M. Hassan



The Electrochemical Society
Advancing solid state & electrochemical science & technology

241st ECS Meeting

May 29 – June 2, 2022 Vancouver • BC • Canada

Abstract submission deadline: Dec 3, 2021

Connect. Engage. Champion. Empower. Accelerate.
We move science forward



Submit your abstract



Feature Selection Using Multivariate Adaptive Regression Splines in Telecommunication Fraud Detection

Muhalim Mohamed Amin^{1,2}, Anazida Zainal^{1,2}, Nurulhuda Firdaus Mohd. Azmi³ and Nor Azizah Ali^{1,4}

¹School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Johor, Malaysia.

²Cyber Threat Intelligence Lab, Information Assurance & Security Research Group, UTM, Johor, Malaysia.

³Razak Faculty of Technology and Informatics, UTM, Kuala Lumpur, Malaysia.

⁴Applied Industrial Analytics Research Group, UTM, Johor, Malaysia.

E-mail: muhalim@utm.my

Abstract. Feature selection determines the most significant features for a given task while rejecting the noisy, irrelevant and redundant features of the dataset that might mislead the classifier. Besides, the technique diminishes the dimensionality of the attribute of the dataset, thus reducing computation time and improving prediction performance. This paper aims to perform a feature selection for classification more accurately with an optimal features subset using *Multivariate Adaptive Regression Splines* (MARS) in Spline Model (SM) classifier. A comparative study of prediction performance was conducted with other classifiers including Decision Tree (DT), Neural Network (NN) and Support Vector Machine (SVM) with similar optimal feature subset produced by MARS. From the results, the MARS technique demonstrated the features reduction up to 87.76% and improved the classification accuracy. Based on the comparative analysis conducted, the Spline classifier shows better performance by achieving the highest accuracy (97.44%) compared to other classifiers.

1. Introduction

Fraud in telecommunication has been a major challenge to the growth of the industry all over the world that suffers major losses due to fraud activities every year. Various types of prediction models had been proposed in the literature to minimize such losses but still facing some limitations that prone to fail, low accuracy rates and high false alarms rates for classification. These are due to the reality that the datasets are getting larger with more attributes to represent the user's behavioral patterns. Instead of depending on the good predictive models, the features selection to get an optimal features subset by reducing the number of features does play an important role to yield better discriminative capability of the models.

In predictive models, the feature selection stage assists in eliminating the irrelevant features from the dataset that primary concern for better accuracy [1]. A larger number of features make more sparse to the dataset and more training data are necessary to accurately sample such a large dataset. Therefore, a good feature selection method is required to speed up the processing, reduces the time and improving

¹ To whom any correspondence should be addressed.



the predictive performance accuracy [2]. The MARS as a wrapper method has been chosen as feature selection because it always provides the best subset of features on many occasions and the model more prone to overfitting with the subset of features compared filter method.

The remainder of this paper starts with the section describing some works that motivated the usage of MARS as the features selection technique in some applications and the MARS modeling itself. It follows by a description of conducted experiments to perform the comparative analysis among four different classifiers (DT, SVM, NN, SM) using an optimal features subset by MARS. Then, the subsequent section addresses the results and discussion before a general conclusion in the final section.

2. Multivariate Adaptive Regression Splines (MARS)

MARS algorithm [3] considered a non-parametric regression modeling procedure. It uses splines to fit piecewise continuous functions to model responses across the entire range of each variable that differently to normal linear regression techniques. This modern statistical learning model performs self-determines in producing a subset of features that best predict a target field of interest, and greater than a conventional logistic regression.

Basically, MARS is capable to identify a relatively small number of predictor variables, which are fairly complex transformations of initial variables. In [4], they investigated MARS together with logistic regression in modeling direct response behavior for direct marketing. They reduced the predictor variables to 15 or only used about 7.5% out of 200 original predictor variables that enhance the predictive capability compared to existing modeling methods. Another research in [5] applied MARS in classifying the Ischemic and Hemorrhagic modified risk factors also improved accuracy by overcoming the missing value. In [6], the study predicted the uplift displacement and evaluate the underground structure floatation in terms of structural characteristics, soil properties, and earthquake parameters. They found that the MARS was demonstrated better accuracy and reliability with the error within approximately $\pm 20\%$. From the viewpoint of mining and civil engineering operations, [7] proposed MARS as a new alternative method to predict blast-induced ground vibration. Their statistical analyses exposed that the MARS demonstrated the best performance in their model compared to Artificial Neural Network (ANN) and conventional ground vibration predictors.

2.1 MARS Modeling

MARS utilizes the advantages of the simplest algorithm in supervised learning, which is linear regression such as it eases and speeds of computation, and also the intuitive nature of interpreting their coefficients. The MARS model uses the form of an expansion in multivariate spline basis functions:

$$Y_i = f(X_i) = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \text{for } i=1,2,\dots,n \quad (1)$$

where Y_i and X_i represent the i -th response value and feature value, respectively, β_0 and β_1 are fixed of coefficients or parameters that represent the intercept and slope of the regression line, respectively, and ε_i represents noise or random error.

The Equation (1) can be extended to capture any non-linear relationship by explicitly including polynomial terms (e.g., X_1^2) or step functions. Polynomial regression is a form of regression in which the relationship between X and Y is modeled as a d -th degree polynomial in X . The polynomial regression function is represented as:

$$y_i = \beta_0 + \beta_1 C_1(x_i) + \beta_2 C_2(x_i) + \beta_3 C_3(x_i) \cdots + \beta_d C_d(x_i) + \varepsilon_i \quad (2)$$

where y is modeled as a d -th degree polynomial in X , $C_1(x)$ represents X values ranging from $c_1 < X < c_2$, $C_2(X)$ represents X values ranging from $c_2 < X < c_3$, \cdots , $C_d(X)$ represents X values ranging from $c_{d-1} < X < c_d$. MARS chooses basis functions for approximating the response through two-stage (*forward* and *backward* stepwise) processes to apply an adaptive regression procedure to produce an optimal MARS in finding the location and number of the needed spline basis functions. Firstly, a very substantial number of basis functions were constructed to overfit the dataset initially. Secondly, the overfitting

spline function for each knot will remove the least contribute knots to the overall fit of the model as determined by the *Generalized Cross-Validation* (GCV) model fit criterion [8] that eliminating the most insignificant variables. The *Lack-of-Fit* (LOF) criterion is used by MARS to evade an excessive number of spline basis functions:

$$LOF(f_M) = GCV(M) = \frac{1}{n} \left(\frac{\sum_{i=1}^n [y_i - f_M(x_i)]^2}{\left[1 - \frac{C(M)}{n}\right]} \right) \quad (3)$$

where

$$C(M) = M + dc \quad (4)$$

In Equation (3) and (4), n denotes as the number of sample observations, $C(M)$ is the number of linearly independent basis functions, M is the number of knots selected in the forward process, while d and c are degrees of interactions and the number of basis functions that consists of spline functions, respectively. A smaller number of knots and smoother function estimates can be achieved by larger values of d . The best MARS approximation is the one with the highest GCV value [9].

3. Experiments

3.1. Data Preprocessing

Like all real-world data, the telecommunication Call Details Record (CDR) dataset contains errors and noise from various sources. The data preprocessing is the first important step that is carried out to remove any type of irregularities, inaccurate or missing values entries that could mislead the pattern of user behavior from the datasets by using the *Alteryx Designer 2019.3* platform. The experiment was run with 5-fold cross-validation with 70% and 30% for training and testing datasets, respectively.

3.2. Features Derivation

The deriving feature is about creating new input features representing the user's behavior usage for the predictive model, and one of the most effective ways to improve predictive models performance [1, 2, 10]. The original CDR that consists of 12 raw features. A total of 58 features were derived that compiled with 23 subset features from literature and 35 additional derived features.

3.3. Features Selection

The MARS has been configured with the basic parameters and executed on each datasets using all 58 features with the Spline model as the classifier. Each dataset produced an optimal features subset in different ranked based on relative importance level. The best optimal features subset was chosen based on the highest overall accuracy, the minimal number of features obtained and low false alarms error. This selected optimal features subset then was applied to all classifiers to do the classification.

3.4. Classification

The predictive model as classifier used in this study was the Splines model as a modern statistical learning model. The performances of this model have been investigated and analyzed with another modern statistical learning model (DT) and two different traditional statistical models (SVM, NN). All related results were recorded for each classifier that executed on each dataset and the performance of overall accuracy was averaged each. All compared classifiers have been configured with a basic parameter, respectively.

4. Results and Discussion

The results of all optimal features subset summarized in Table 1. Based on the table, dataset 4 reduced the lowest number of features with the highest reduction percentage of 84.48%, while the rest obtained 82.76% of reduction. Different instances with various features values been distributed randomly in each dataset might demonstrate this result. However, the best optimal feature subset was selected from dataset 1 that demonstrated the highest average overall accuracy of 97.81% with the FP and FN rates are 1.05% and 4.78%, respectively. This optimal feature subset has been used for all classifiers to identify which one shows better performance.

Table 1. Features subset for each dataset

	Dataset				
	1	2	3	4	5
Optimal Features	10	10	10	9	10
Reduction (%)	0.8276	0.8276	0.8276	0.8448	0.8276
Accuracy (%)	0.9781	0.9729	0.9688	0.9698	0.9708
FP Rate (%)	0.0105	0.0195	0.0180	0.0165	0.0150
FN Rate (%)	0.0478	0.0444	0.0614	0.0614	0.0614

In terms of classification, Table 2 illustrates the highest average correct classification rate is 97.44% performed by the Spline model, followed by the NN model with slightly lower at 96.90%. These two models demonstrated high average accuracy compared to the rests of models because of their flexibility to adapt to more complex limit state functions that might not be represented well employing a low order polynomial. The rest of the models performed with an average of the overall accuracy of approximately less than 96%. Even the DT model required less effort for data preparation during pre-processing, it is inadequate for applying regression and predicting continuous values that could lead a bad results with low accuracy.

Table 2. Average performance of classification

Models	Accuracy	G Mean	Error Rate	FP Rate	FN Rate
DT	0.9429	0.9238	0.0571	0.0291	0.1208
SVM	0.9569	0.9413	0.0431	0.0201	0.0956
NN	0.9690	0.9630	0.0310	0.0219	0.0519
Spline Model	0.9744	0.9662	0.0256	0.0132	0.0539

Since the class or target in datasets was not extremely imbalanced with an acceptable proportion of 70% and 30% for the normal and fraud samples [11], respectively, the Geometric Means (GM) accuracy can also be used to evaluate the performance of each model. As can be seen through Table 2, each model demonstrated the ranking level of average GM accuracy remained similar to the average global accuracy, respectively, with the highest is 96.62% performed by the Spline model.

From the perspective of misclassification error, it can be observed from Table 2 that the lowest average percentage is 2.56% performed by Spline model with scores of 1.32% and 5.39% for FP and FN rates, respectively. It followed by the NN, SVM and DT models with the average percentage of error rates were approximately 3%, 4%, and 6%, respectively. In fraud detection, misclassification costs for both FP and FN rates are unequal, uncertain, can differ from dataset to dataset, and can change over time. Besides, the latter error is usually more costly than a first error [12]. In general, the NN and Spline models demonstrated the FN errors of less than 6% while the rest of the models performed a little bit higher with more than 9%. The DT model indicates a higher FN error rate at 12% might be due to its high probability of overfitting that gives low classification for a dataset. For the SVM model with approximately 10% of the FN error rate, it has generalization in practice to reduce the risk of overfitting,

but it is not suitable for large datasets even it works relatively well with a clear margin of separation between classes.

5. Conclusion

Feature selection is very significant since the datasets granted for a research investigation may comprise thousands of features with some features that might be irrelevant. In this paper, an optimal feature subset was presented using the MARS approach as the feature selection approaches that were applied to four various types of models or classifiers. MARS was executed on 5 different datasets in the conducted experiment and the best optimal features with high overall accuracy and low false alarm rate were chosen to be used in developing a classification model. The results demonstrated that MARS conserves classification accuracy i.e. unnecessary features can be eliminated efficiently from the datasets without decrease the classification performance. From this study, the feature selection methods using MARS show better reduction by using only approximately 17% or 10 selected features with high relative importance and the Spline classifier demonstrated high classification accuracy outperforms other classifiers. In conclusion, choosing the right features can produce simpler and more flexible models to yield better results of prediction by a classifier. The potential of the MARS approach as a feature selection method can be used in other applications of fraud detection that concern a better overall accuracy.

References

- [1] Y. Masoudi-Sobhanzadeh, H. Motieghader, and A. Masoudi-Nejad. FeatureSelect: A Software for Feature Selection Based on Machine Learning Appr. *BMC Bioinformatics*, **20** (170), 2019.
- [2] K. J. Mathai and K. Agnihotri. Optimization Techniques for Feature Selection in Classification. *International Journal of Engineering Development and Research*, **5** (3):1167–1170, 2017.
- [3] J. H. Friedman and C. B. Roosen. An Introduction to Multivariate Adaptive Regression Splines. *Statistical Methods in Medical Research*, **4** (3):197–217, 1995.
- [4] J. Deichmann, A. Eshghi, D. Haughton, S. Sayek, and N. Teebagy. Application of Multiple Adaptive Regression Splines (MARS) in Direct Response Modeling. *Journal of Risk and Insurance*, **16** (4):15 – 27, 2002.
- [5] R. D. L. N. Karisma and S. Harini. Multivariate Adaptive Regression Spline in Ischemic and Hemorrhagic Patient (Case Study). In *AIP Conference Proceedings*, **2084**, Mac 2018.
- [6] G. Zheng, P. Yang, H Zhou, C. Zeng, X. Yang, X. He, and X. Yu. Evaluation of the Earthquake Induced Uplift Displacement of Tunnels using Multivariate Adaptive Regression Splines. *Computers and Geotechnics*, **113**:1–10, 2019.
- [7] C. K. Arthur, V. A. Temeng, and Y. Y Ziggah. Multivariate Adaptive Regression Splines (MARS) Approach to Blast-Induced Ground Vibration Prediction. *International Journal of Mining, Reclamation and Environment*, **34** (3):198–222, 2020.
- [8] P. Craven and G. Wahba. Smoothing Noisy data with Spline Functions. Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation. *Numerische Mathematik*, **31**:377–403, 1979.
- [9] S. Crino. Global Optimization With Multivariate Adaptive Regression Splines. *IEEE Transaction on Systems, Man and Cybernetics*, **37**(2):333 – 340, 2007.
- [10] S. Visa, B. Ramsay, A. Ralescu, and E. Knaap. Confusion Matrix-based Feature Selection. In *22th Midwest Artificial Intelligence and Cognitive Science*, 2011.
- [11] B. Baesens, V. V. Vlasselaer, and W. Verbeke. *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection*. John Wiley & Sons, 2015.

- [12] S.N. Shivakumar and S.C. Lingareddy. Fraud Detection using Data Mining Techniques. *International Journal of Innovations in Engineering and Technology (IJJET)*, **4**(1):304– 312, 2014.