Utilizing Data Analytics to Analyze Online Purchase Behavior

David Marshall
Department of Marketing, Eastern Michigan University, Ypsilanti, Michigan, USA
dmarsh12@emich.edu

Received Jan 31, 2023, **Accepted:** Feb 14, 2023, **Published Online:** Feb 28, 2023
Reviewers: Anonymous Peer Review
Citation: Marshall, D. (2023). Utilizing Data Analytics to Analyze Online Purchase Behavior. *International Journal of Supply Chain Management, 12*(1), 45-57, https://doi.org/10.59160/ijscm.v12i1.6164

***Abstract*— The emergence of data analytics has fundamentally transformed supply chain management strategies in the global marketplace during the past decade. Classification is one of the most popular methods and receives a great deal of attention in the literature, but there are still some questions concerning the performance characteristics of different classification methods. This paper analyzes three different classification methods: classification trees, k-nearest neighbors, and artificial neural networks to determine if there are any performance gaps between the methods. A series of experiments are conducted utilizing the Analytic Solver Data Mining (formerly XLMiner) add-in to Microsoft Excel in an effort to address these issues. The analysis reveals that there may be minor performance gaps, but the methods all perform well in the context of this study. The findings indicate that decision trees may be the preferred classification techniques due to the overall performance of the method, ease of use, and an abundance of baseline algorithms that can be employed.**

***Keywords*—** *Supply Chain Management, Knowledge, Competencies, Skills, O*NET*

1. Introduction

The digital revolution of the past couple decades has propelled organizations to collect massive amounts of data in an effort to capture valuable information about the company, customers & suppliers, and competitors, primarily because the data is easy to collect and store [1]. Wide-spread use of information systems and technology facilitates the collection and processing of this data. One example of large-scale data collection can be found in Wal-Mart's teradata-based warehouse that contains well over 1 petabyte (1,000,000,000,000,000 bytes) of sales information generated from 800 million transactions each day [2]. Supply chain analytics has emerged as one of the driving forces of organizational decisions making [3]. Analytics allow supply chains to communicate sales data, forecasts, production scheduling, and a variety of other metrics [4], [5]. The advantages of collecting this volume of data are pretty obvious, but there are several challenges associated with the large-scale datasets. One issue that organizations continue to grapple is the extraction of valuable knowledge from the data for use in the decision making activities within the organization [6]. Organizations also struggle with data privacy and security measures associated with such large datasets [7].

Traditional methods of data analysis relied on experienced statisticians to extract the useful information from a large dataset using a variety of statistical methods, but another technique has emerged in the past few years that can make data analysis more efficient and effective [1]. Data Analytics has materialized into a powerful tool that relies on complex algorithms, in addition to conventional data analysis methods, to process and identify patterns in large datasets [6]. There are many different data analytics techniques in use today that focus on outcomes such as: prediction, classification, cluster analysis, anomaly detection, association analysis, and numerous others. One data analytics approach that has received considerable attention in the literature for an extended period of time is classification (see [8]-[10]). Most publications provide an in-depth analysis of one or more of the classification techniques like decision/classification tree [11]-[14]; nearest neighbor [15]-[17]; and artificial neural network [18], [19], but there are very few studies that compare the performance of different classifiers.

IJSCM, ISSN: 2050-7399 (Online), 2051-3771 (Print)

The purpose of this research is to comparatively investigate the performance of three different classification methods within the context of online retailing in the supply chain. As classification techniques continue to evolve and become increasingly utilized, this study builds upon prior research that analyzed different classification methods including their applicability, data requirements, and ease of use, among other factors [20]. Decision trees, nearest neighbor, and artificial neural network classification schemes represent the foundation for this comparative study because there are similarities between each method yet they also have unique attributes in terms of inputs, outputs, and variable types. A group of experiments are conducted for each method in order to identify key characteristics and features.

2. Literature Review

Data analytics has attracted a tremendous amount of attention in the literature during the past several years. Organizations are increasingly leveraging the massive amounts of data in order to drive organizational decision making, specifically as it relates to supply chain management. Many studies identify and develop analytical techniques for processing data from retailers, however a newer stream of research develops data mining tools in the transportation sector [21], [22]. Classification is a common data mining technique that is widely utilized and relatively simple for organizations to employ, thereby making it a preferred choice for data analysis. A series of seminal and more recent studies highlight the applicability of classification as a tool to analyze organizational data that drive firm performance and sustainable competitive advantage [23]-[25].

Classification can be defined as "the task of assigning objects to one of several predefined categories [6]." Each of the classification methods presented follows a formal iterative learning routine where a training set is used to train the model and the subsequent class labels assigned to each test record are compared to the known class label to generate an error rate for the classification. The following sections provide a brief overview and literature review of the three classification techniques utilized in this research.

A Decision Tree is a formalized classification technique that is constructed by asking a series of questions about the characteristics of a dataset in an effort generate classification rules that can be applied to future data points [6]. Decision trees are based on inductive reasoning, where a set of observed cases are used to construct generalized rules that stem from the observed cases, which are then used to predict or classify future observations [26]. Rule induction techniques are typically used for classification when the dependent variable is categorical/nominal, but the independent variable(s) can be either nominal or interval [27]. For the purpose of this analysis, we consider the terms decision tree and classification tree to be synonymous.

Decision trees are constructed, with the support of complex algorithms, by partitioning the data from a root node (no incoming edges and zero or more out-going edges) into internal nodes (one incoming edge and two or more out-going edges) and leaf nodes (one incoming edge and no out-going edges) based on an iterative process to find the best split that minimizes variance [6]. The leaf node represents the class label for the dependent variable and the internal or root nodes represent attributes of an independent variable. There are many different decision tree algorithms in use today and nearly all of them trace their roots to Hunt's Algorithm [28]. Some of the more popular decision tree algorithms include CART [12], C4.5 [29], CHAID [30] and ID3 [11]. A sample decision tree with arrows that highlight the root, internal, and leaf nodes can be found in Figure 1.

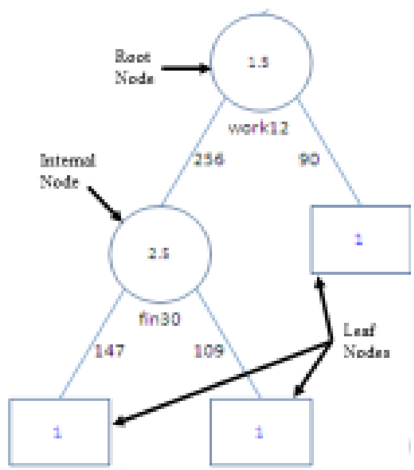
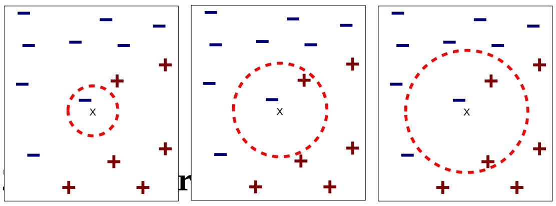


Figure 1. Example Decision Tree

According to ref [6], the sample decision tree depicted above is an example of an 'eager learner' because "it is designed to learn a model that maps the input attributes to the class label as soon as the training data becomes available", whereas an artificial neural network is considered a 'lazy learner' because "the process of modeling the training data is delayed until it is needed to classify the test examples (p. 223)". Nearest neighbor classification is accomplished by classifying unknown data points based on points that have similar attributes in close proximity to the unknown data point [6]. A k-nearest neighbor (kNN) classification algorithm attempts to find the k neighbors nearest to an unknown data point based on a proximity measurement (usually Euclidean distance) in an effort to classify the unknown point with the same class label as its k nearest neighbors, where k represents the number of data points used for the comparison [31]. A popular quote to summarize k-nearest neighbor classification is: "If it walks like a duck, quacks like a duck, and looks like a duck, then it's probably a duck [6](p. 224)". The selection of k is extremely important due to the potential issues of overfitting and misclassification. Figure 2 displays the 1-, 2-, and 3-nearest neighbor classification structure of a data point x located at the center of each circle, where the classification of point x is chosen based on the majority class of its k-nearest neighbors. As you can see from the figure, the selection of k=1 would classify the point as negative (-), the selection of k=2 would be a tie and the class label would be selected randomly, and the selection of k=3 would classify the point as positive (+).



(a) 1-nearest neighbor (b) 2-nearest neighbor (c) 3-nearest neighbor

Figure 2. The 1-, 2-, and 3-nearest neighbors of point x [6]

The artificial neural network (ANN) classification technique traces its roots to biological neural systems, namely the neural structure of the human brain [32]. An artificial neural network consists of an input layer which contains the independent variable(s), a hidden layer(s) with embedded (hidden) nodes, and an output layer that contains the dependent variable(s). There may be multiple hidden layers in an artificial neural network. Nodes in each layer are connected by arcs and signals are transmitted from the independent variables in the input layer, forward through the hidden layer, to the dependent variable in the output layer via the connecting arcs [32]. The artificial neural network is constructed by specifying the number of layers and the number of nodes in each layer and then the training data with known values for the class label is fed to create the network structure. Once the neural network is constructed, it can then be pruned to simplify the model and remove redundant links and/or nodes as long as the classification error rate does not significantly increase [33].

There are many different applications where ANNs can be used as a means of classification including: cancer prediction, fault detection, security/intrusion, and ecological modeling, to name a few. Some authors argue that artificial neural networks have many flaws such as: long learning time, difficult to specify, or the difficulty associated with extracting knowledge from the ANN [28]. Another issue with ANNs is that the hidden layer is essentially a black box with an embedded classification scheme that makes deciphering the model extremely complicated. Figure 3 illustrates a typical artificial neural network with an input layer containing five independent variables, one hidden layer, and an output layer with one dependent variable.

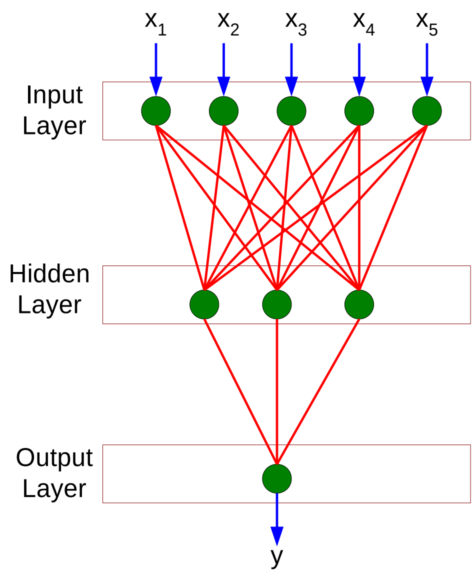


Figure 3. Example of a multilayer ANN [6]

3. Research Methodology

The methodology employed in this study centers on the data analytics classification techniques of decision trees, k-nearest neighbors, and artificial neural networks. The following sections offer a brief description of the data collection procedures, pre-processing efforts, and an overview of each classification method. The Analytic Solver Data Mining (formerly XLMiner) add-in for Microsoft Excel is the data analytics software package utilized for this research, herein referred to as Solver, which is primarily due to the ease of use and availability of the product.

3.1. Data Collection

The National Opinion Research Center (NORC) was founded in 1941 as a social science research institution with support from private corporations, educational institutions, not-for-profit organizations, and the U.S. government [34]. The headquarters for NORC is located at The University of Chicago, and it has regional offices that stretch from Washington D.C. to California. NORC conducts interdisciplinary social science research on a variety of topics utilizing many different research techniques. The General Social Survey (GSS) is one of NORC's oldest projects and serves as the data source for this study. The GSS, which originated in 1972 and is funded by the sociology branch of the national science foundation, is a compilation of suggested research questions from researchers around the country. Most of the questions from the first survey in 1972 are still used today in order to track research trends over time. According to NORC's website, the GSS is the only survey that has documented the opinions of Americans over such a long time period and is the most frequently analyzed source of social science data excluding the U.S. Census [34].

The general social survey data is available to download from the GSS website for use with three statistical packages: SPSS, STATA, and SAS. The entire dataset collected since 1972 can be downloaded in a single file, or data can be selected from any year that the survey was administered. For the purpose of this research, the cumulative data file from 1972-present was collected in an effort to maximize the amount of available data and observations. The initial data file contains approximately 5,100 variables and nearly 51,000 observations. The SPSS software package is the selected repository for data delivery.

3.2. Data Pre-processing and Sampling

Data analytics techniques work very well on large datasets, but many techniques can only handle the depth of data with minimal tolerance for data breadth. The classification techniques explored in this research with Solver are limited to only 200 variables and 60,000 observations, so the dataset requires extensive pre-processing in order to prep the data for the analysis. Due to the nature of the cumulative data file, there is an excessive amount of missing data that must be purged from the dataset, which primarily stems from the issue that there are roughly 3500 variables in the dataset that only contain data for one year of the survey. This abundance of missing data further amplifies the need for variable reduction.

Table 1 presents the sample composition for online purchase behavior. As you can see in the table, the original dataset contained 5084 variables and 51020 observations. The first step of sample selection involves deleting the observations with missing data for the dependent variable – online purchasing habits. Since the data for online purchasing was only collected in the years 2000 and 2002, all observations prior to the year 2000 and after the year 2002 contained missing data for the online purchase variable and were subsequently deleted from the analysis. This reduced the number of observations from 51020 to 1655. The next step of the process dealt with missing data in the remaining variables. Variables that contained greater than 10% missing data (166 or more observations missing) during the 2000-2002 window were deleted from the dataset, which reduced the variables in the analysis from 5084 to 61. Summary statistics were generated for the remaining 61 variables, and the mode for each variable was used to replace missing data within the 1655 remaining observations for that variable. The final step of the sample selection considered data that was conceptually irrelevant to the analysis of online purchase behavior. For example, an individual's astrological sign was considered to be conceptually irrelevant to a decision to make an online purchase. The final sample for the study contains 49 variables with 1655 observations.

Table 1. Online Purchase Sample Composition

Description	Variables	Observations
Original data	5084	51020
Less:		
Unavailable data	-	(49365)
Missing data > 10%	(5023)	-
Conceptually irrelevant data (e.g. astrological sign)	(12)	-
Online purchasing behavior sample	49	1655

There are two distinct categories that emerge for the remaining 49 variables used in the analysis. The first category belongs to variables that pertain to web browsing history. The web browsing variables relate to web sites visited during the past 30 days (xxxx30) and web sites visited during the past twelve months (xxxx12). The dependent variable for this study is a nominal variable with response categories of "yes" & "no" that is based on the question "Have you made an online purchase in the past twelve months?."

Solver also has limitations on the number of inputs used, which is strictly dependent upon the technique used and variable type. Categorical variables greatly inhibit the inputs for a given method because Solver requires that categorical variables be converted to m-1 binary variables, where m represents the number of categories in the variable. Based on these limitations, some variables were re-coded in an effort to streamline the categories within the variable and to address the input restrictions imposed by Solver. For example, the variable 'region' originally contained nine categories (New England, Middle Atlantic, E. North Central, W. North Central, South Atlantic, E. South Central, W. South Central, Mountain, Pacific) and was reduced to four classes based on the time zone (Eastern, Central, Mountain, Pacific). Finally, the data is partitioned into a training set that contains 50% of the observations, a validation set containing 30%, and a test set with 20% of the data.

3.3. Experimental Design

The Solver add-in has limitations on the number and type of input variables accepted for each method. The decision tree, k-nearest neighbor, and artificial neural network classification techniques can all accept continuous and ordinal input variables but not nominal variables. To address the issue with nominal variables, the data utilities menu of Solver can transform a nominal variable with m categories into m-1 binary variables, as mentioned above. Solver has a limitation of 10,000 training observations and 60,000 total observations including the training, validation, and test sets, which does not impact this study since the dataset contains only 1655 observations. The number of inputs, which includes the binary variables generated from the nominal variables, is limited to 30. Due to the limit on input variables, the dataset is divided into two subsets. One subset contains the variables that correspond to online browsing history and the other subset includes demographic variables that pertain to the respondents (i.e. marital status). There are also individual limitations for each of the three methods. The number of k nearest neighbors for the kNN classifier is limited to twenty and the maximum number of levels displayed in the classification tree is seven. The primary limitation for the output is that Solver cannot handle continuous dependent variable, but it can accept a nominal dependent variable, which is desirable for the classification methods employed in this research. If it was not explicitly mentioned above, it is assumed that any other limitations imposed by Solver do not affect this analysis.

To facilitate a thorough comparison of the three different classification techniques, a series of experiments is conducted for each of the three methods. To begin the experiment, two different partitioning methods are used with 50% of the data assigned to the training set, 30% assigned to the validation set, and 20% assigned to the test set. The next step is to evaluate each classification method to understand the difference between normalized and non-normalized data. Under the umbrella of normalized and non-normalized data, each method is replicated five times to capture the average performance and variance in performance. Table 2 presents an experimental design matrix for this study. Each experiment is replicated 5 times to capture performance statistics.

Table 2. Design of Experiments

	Web Usage Variables	Demographic Variables	Combination of Web Usage and Demographic Variables
Standard Partitioning			
With data normalization	*	*	*
Without data norm.	*	*	*
Partitioning with Oversampling			
With data normalization	*	*	*
Without data norm.	*	*	*

4. Results

The primary objective of classification is to create a model that can accurately predict the most likely class of an unknown record with the smallest possible misclassification (error) rate. There are many different parameters that can be manipulated in an effort to achieve the lowest possible error rate. The purpose of this research is to compare the performance of three different classification techniques, but the experiments employed for the evaluation of the different methods are not designed to minimize the overall error rate. Therefore, the misclassification error rate will be the basis for the comparisons between the different methods and there is no effort to optimize or improve the error rate within the individual techniques.

4.1. Classification Tree

After some investigation, it was determined that partitioning with oversampling is not an ideal experimental setting for this research because that particular partitioning method is typically used when the percentage of the success category is very low (less than 5-10%). The success category for this research is a 'yes' response to the question assigned as the dependent variable, which accounts for approximately 58% of the total responses.

The results for the classification tree experiments are displayed in a table in Appendix A below. The default parameters were used for the analysis with the exception of the normalization option. Detailed reports were generated for the training, validation, and test sets. Each routine was replicated five times in order to capture the average and variance in performance. There was no noticeable difference in performance based on the output of the replicated experiments. This may be attributed to the fact that the same parameters were used for each replication.

As you can see from the table, there was no difference in performance for the normalized and non-normalized data for any of the trials. According to the help section in Solver, normalizing the data will only produce different results if linear combinations of the inputs are used for splitting. The overall error rates fall within the range of 31-36%, which seem high for this model, but again, the purpose of the study is not to minimize error rates. To form a consistent boundary for the comparison, the test set is used as the basis for model judgement. The model that contains the combination of web usage and demographic variables outperforms the models that contain only the web usage or demographic variables alone. Even though the overall error rate for the web usage variables is the same as the overall error rate for the combination of variables, the error rate for the success category (Yes) is much lower for the model with the variable combinations and provides justification for selection of this model.

The final classification tree is shown in Figure 4 below. The root node is formed with the variable 'buyinf12'. The value inside the circle (node) is the split threshold and can be interpreted as follows: records with buyinf12 <= 1.5 follow the left branch and records with buyinf12 > 1.5 follow the right branch. The numbers beside the branches identify the number of records that follow either branch. In other words, the 164 records that follow the right branch will be classified as a "no" class label for the dependent variables and the 664 records that follow the left branch are further split by the internal node that represents the variable 'wwwmin'. This pattern continues until all records are classified at a terminal or leaf node, and the internal nodes are interpreted in the same manner

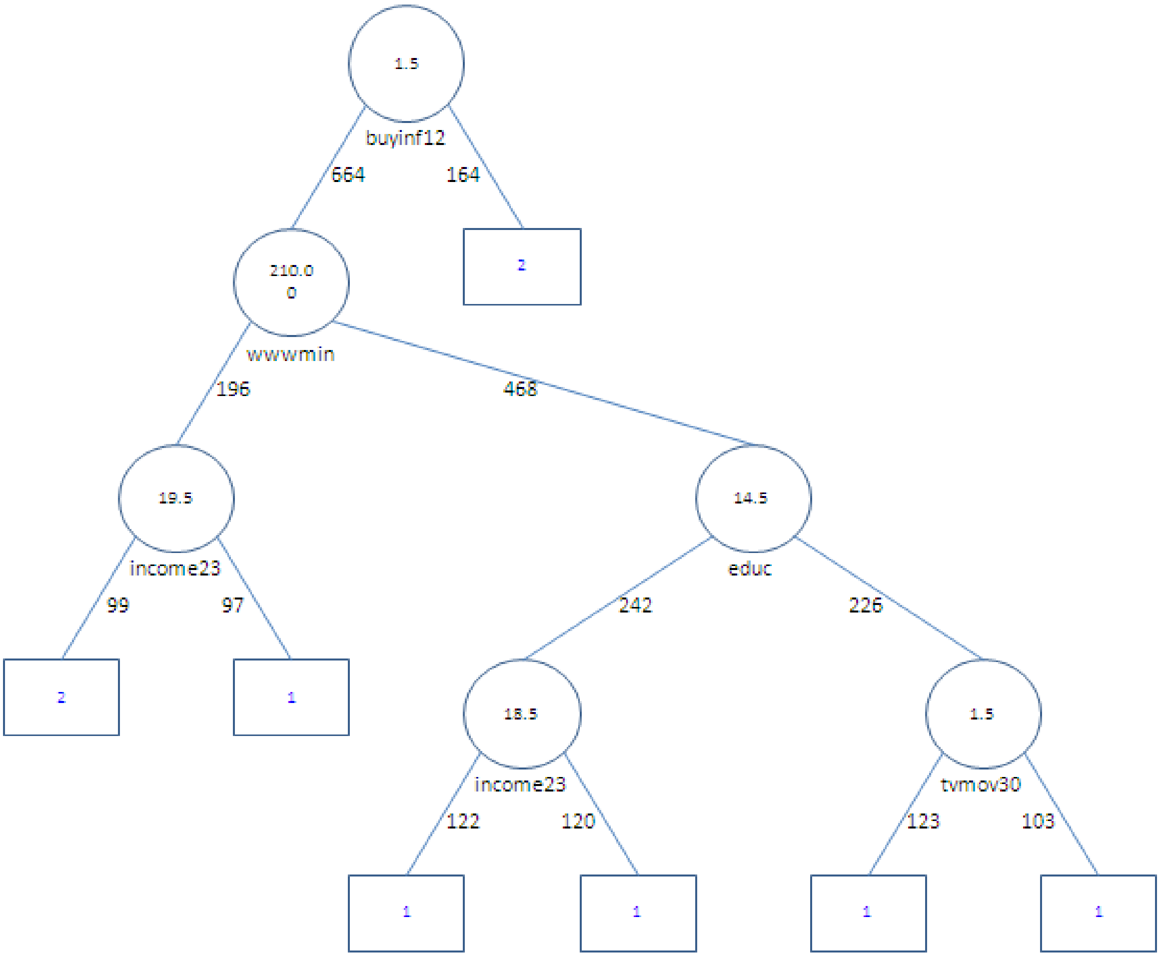


Figure 4. Full Classification Tree for the Combination of Web and Demographic Variables

as the root node described above. The full classification tree for this analysis is shown below for visual purposes, but the tree can be pruned in order to simplify the rules generated by the tree.

4.2. k-Nearest-Neighbor

Similar to the classification tree method, the default parameters were used for the k-nearest neighbor technique with the exception of one key difference. Solver limits the number of k-nearest neighbors at twenty. There is also an option that can be selected during model specification that will enable the software to select the best value for k based on the lowest error rate for the validation set. This option was selected for this analysis and the k-values vary depending on the experiment. Table 3 presents a summary of the best k selection process with the corresponding error rates listed for each value of k up to twenty. The best value for k listed in the table is associated with the experiment for the combination of web usage and demographic variables with normalized data. The experiments for the web usage variables were replicated five times for both the normalized data and non-normalized data. The results of the kNN replication are the same as the results of the classification tree replication – no observable difference in performance.

Appendix B below contains the error rates for k-nearest neighbor trials. In this case there was a difference in performance for the normalized and non-normalized data. Standardization of the variables is important for the kNN method because large differences in the scale of categorical variables can influence the distance measure. Again, the test set is used to compare the performance, but in this case it is tough to determine which model between the combination of web usage and demographic variables and the web usage variables is best performer. The overall error rate is virtually the same for both experiments, but the error rate for the success category (Yes) is slightly lower for the model with the variable combinations.

Table 3. Error Log for the Selection of the Best k-Value

Value of k	% Error Training	% Error Validation	
1	0.00	37.30	
2	20.89	37.90	
3	18.60	35.28	
4	22.71	34.88	
5	23.31	32.86	
6	25.72	34.27	
7	24.28	34.27	
8	27.05	35.48	
9	26.33	35.48	
10	27.90	34.68	
11	27.05	32.66	
12	27.66	33.06	
13	28.14	32.26	
14	27.78	31.85	
15	27.29	30.65	
16	27.78	31.25	
17	27.66	30.24	<---- Best k
18	28.86	30.65	
19	28.74	30.44	
20	28.86	31.25	

4.3. Artificial Neural Network (ANN)

An additional consideration for the ANN classification method that was not required for the two prior methods is the number of hidden nodes to include in the model. Solver will permit from 1-4 hidden layers to be included in the network. The first step of the ANN classification included an evaluation of the performance based on hidden layers from 1-4 and concluded that a hidden layer = 2 is the best design. The default parameters were also used for the artificial neural network experiments.

The misclassification error rates for the ANN method are presented in Appendix C below. Similar to the kNN results, the misclassification error rates are different for the normalized vs. non-normalized data. The error rates for the non-normalized data with the demographic and combination variables are extremely large, which suggests that there is tremendous influence by differences in the scale for the variables or that there may be an unforeseen issue with the analysis. For the ANN analysis, the web usage variables with normalized data outperform the other two groups, both in terms of overall error rate and error rate of the success category.

5. Conclusion

The performance of each classification technique in terms of the overall error rates is comparable. There are distinct advantages and disadvantages to each method. The performance of the classification tree is slightly better than the other two, so the ideal classification method for the data in this study is classification tree. Despite some of the limitations, classification trees are relatively easy to understand and the rules generated from the tree are straightforward and useful. Many organizations develop proprietary tools for analyzing data; however, this study provides support for classification trees as a viable alternative.

This research has two primary contributions. First, organizations have realized the value of leveraging data analytics for decision making, and this research applies a relatively simplistic approach to data analysis that can easily be replicated in industry. Second, an evaluation of the performance characteristics of the different classification techniques lends additional insight to supply chain analytics techniques, and provides a pathway for future research.

References

[1] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. *"The KDD process for extracting useful knowledge from volumes of data,"* Communications of the ACM, Vol 39, No. 11, 27-34, 1996.
[2] Schell, E., HP Data Warehouse at Wal-Mart. Direct Commerce Systems and Services. Retrieved October 10, 2022, from http://directcommercesystems.blogspot.com/2007/08/hp-data-warehouse-at-wal-mart.html
[3] Souza, G. C., *"Supply chain analytics,"* Business Horizons, Vol 57, No. 5, 595-605, 2014.
[4] Zhu, S., Song, J., Hazen, B. T., Lee, K., & Cegielski, C., *"How supply chain analytics enables operational supply chain transparency: An organizational information processing theory perspective,"* International Journal of Physical Distribution & Logistics Management, Vol 48, No. 1, 47-68, 2018.
[5] Srinivasan, R., & Swink, M., *"An investigation of visibility and flexibility as complements to supply chain analytics: An*

organizational information processing theory perspective," Production and Operations Management, Vol 27, No. 10, 1849-1867, 2018.

[6] Tan, P. N., Steinbach, M., & Kumar, V., *Introduction to data mining*. Pearson Education India, 2016.

[7] Ogbuke, N. J., Yusuf, Y. Y., Dharma, K., & Mercangoz, B. A., *"Big data supply chain analytics: ethical, privacy and security challenges posed to business, industries and society,"* Production Planning & Control, Vol 33, No. 2-3, 123-137, 2022.

[8] Michie, D., Spiegelhalter, D.J., and Taylor, C.C., "*Machine Learning, Neural and Statistical Classification,"* Ellis Horwood, Upper Saddle River, NJ, 1994.

[9] Cherkassky, V., and Mulier, F, "*Learning from Data: Concepts, Theory, and Methods,"* Wiley Interscience, 1998.

[10] Hastie, T., and Tibshirani, R, *"Classification by pairwise coupling,"* Annals of Statistics, Vol 26, No. 2, 451-471, 1998.

[11] Quinlan, J. R., *"Discovering rules by induction from large collections of examples,"* Expert Systems in the Micro Electronic Age, ed. D. Michie. The Edinburgh University Press, Edinburgh, 1979.

[12] Breiman, L., Friedman, J.H., Olshen, R., and Stone C.J., "*Classification and Regression Trees,"* Chapman & Hall, New York, NY, 1984.

[13] Buntine, W., *"Learning classification trees,"* Artificial Intelligence Frontiers in Statistics, Chapman & Hall, London, 182-201, 1993.

[14] Murthy, S.K., *"Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey,"* Data Analytics and Knowledge Discovery, Vol 2, No. 4, 345-389, 1998.

[15] Cost, S., and Salzberg, S., *"A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features,"* Machine Learning, Vol 10, No. 1, 57-78, 1993.

[16] Cover, T.M., and Hart, P.E., *"Nearest Neighbor Pattern Classification,"* Knowledge Based Systems, Vol 8, No. 6, 373-389, 1995.

[17] Han, E.-H., Karypis, G., and Kumar, V., *"Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification,"* Proc. of the 5th Pacific-Asia Conference on Knowledge Discovery and Data Analytics, Lyon, France, 2001.

[18] Andrews, R., Diederich, J., and Tickle, A., *"A Survey and Critique of Techniques For Extracting Rules From Trained Artificial Neural Networks,"* Knowledge Based Systems, Vol 8, No. 6, 373-389, 1995.

[19] Bishop, C.M., "*Neural Networks for Pattern Recognition,"* Oxford University Press, Oxford, U.K., 1995.

[20] Sulaiman, M. A., *"Evaluating data mining classification methods performance in Internet of things applications,"* Journal of Soft Computing and Data Mining Vol 1, no. 2, 11-25, 2020.

[21] Elazeem, A., Mahmoud, N., Labib, N.M., & Abdella, A.K., *"A proposed data warehouse framework to enhance decisions of distribution system in pharmaceutical sector,"* Egyptian Computer Science Journal 43, no. 2 (2019): 43-60.

[22] Islam, S.B., and Habib, M.M., "Applying Data Mining Tools in Transportation: Data-Driven Supply Chain View." Int. J Sup. Chain. Mgt Vol 10, no. 1 (2021): 29.

[23] Ngai, E. WT, Xiu, L. & Chau, D CK., *"Application of data mining techniques in customer relationship management: A literature review and classification,"* Expert systems with applications Vol 36, No. 2, 2592-2602, 2009.

[24] Zaini, B.J., and Mahmuddin, M., *"Classifying firms' performance using data mining approaches,"* International Journal Supply Chain Management Vol 8, No. 1 690-696, 2019.

[25] Gorade, S.M., Deo, A., & Purohit, P., *"A study of some data mining classification techniques,"* International Research Journal of Engineering and Technology Vol 4, No. 4, 3112-3115, 2017.

[26] Holland, J. H., Holyoak, K. J., Nisbett, R. E. and Thagard, P. R., "*Induction: Process of Inference, Learning, and Discovery,"* The MIT Press, Cambridge, MA, 1986.

[27] Mingers, J., *"Rule induction with statistical data: a comparison with multiple regression,"* Journal of the Operational Research Society, Vol 38, No. 4, 347-351, 1987.

[28] Hunt, E., Marin, J. and Stone P., "*Experiments in Induction,"* Academic Press, New York, NY, 1966.

[29] Quinlan, J.R., *"C4.5: Programs for Machine Learning,"* Morgan-Kaufmann Publishers, San Mateo, CA, 1993.

[30] Kass, G.V., *"An Exploratory Technique for Investigating Large Quantities of Categorical Data,"* Applied Statistics, Vol 29, No. 2, 119-127, 1980.

[31] Khan, M., Ding, Q., and Perrizo, W., Lecture Notes in Computer Science. cs.ndsu.nodak.edu, 2002.

[32] Spangler, W.E., May, J.H., and Vargas, L.G., *"Choosing Data-Analytics Methods for Multiple Classification: Representational and Performance Measurement Implications*

for Decision Support," Journal of Management Information Systems, Vol 16, No. 1, 37-62, 1999.
[33] Lu, H., Setiono, R., and Liu, H., *"Effective Data Analytics Using Neural Networks,"* IEEE Transactions on Knowledge and Data Engineering, Vol 8, No. 6, 957-961, 1996.
[34] National Opinion Research Center, University of Chicago, Retrieved October 23, 2022, from http://www.norc.org/GSS+Website/About+GSS/About+NORC/.

Appendix A. Error Rates for Classification Tree Technique

Classification Tree		Class Label	Standard Partitioning	
			With Data Normalization	Without Data Normalization
			% Error	% Error
Web Usage Variables	Training Set (828)	Yes (1)	14.67	14.67
		No (2)	50.87	50.87
		Overall Error	29.71	29.71
	Validation Set (496)	Yes (1)	40.49	40.49
		No (2)	21.70	21.70
		Overall Error	32.46	32.46
	Test Set (331)	Yes (1)	37.50	37.50
		No (2)	22.30	22.30
		Overall Error	31.12	31.12
Demographic Variables	Training Set (828)	Yes (1)	17.98	17.98
		No (2)	55.23	55.23
		Overall Error	33.45	33.45
	Validation Set (496)	Yes (1)	26.06	26.06
		No (2)	52.36	52.36
		Overall Error	37.30	37.30
	Test Set (331)	Yes (1)	31.25	31.25
		No (2)	42.45	42.45
		Overall Error	35.95	35.95
Combination of Web Usage and Demographic Variables	Training Set (828)	Yes (1)	15.29	15.29
		No (2)	45.06	45.06
		Overall Error	27.66	27.66
	Validation Set (496)	Yes (1)	16.55	16.55
		No (2)	41.98	41.98
		Overall Error	27.42	27.42
	Test Set (331)	Yes (1)	17.71	17.71
		No (2)	49.64	49.64
		Overall Error	31.12	31.12

Appendix B. Error Rates for k-Nearest Neighbor Technique

k-Nearest Neighbor		Class Label	Standard Partitioning	
			With Data Normalization	Without Data Normalization
			% Error	% Error
Web Usage Variables (k=18 for normalized data & k=19 for non-normalized data)	Training Set (828)	Yes (1)	18.80	25.21
		No (2)	46.22	41.28
		Overall Error	30.19	31.88
	Validation Set (496)	Yes (1)	17.25	26.76
		No (2)	46.70	41.04
		Overall Error	29.84	32.86
	Test Set (331)	Yes (1)	17.19	23.44
		No (2)	52.52	42.45
		Overall Error	32.02	31.42
Demographic Variables (k=15 for normalized data & k=9 for non-normalized data)	Training Set (828)	Yes (1)	11.98	18.80
		No (2)	56.40	48.26
		Overall Error	30.43	31.04
	Validation Set (496)	Yes (1)	13.03	24.30
		No (2)	71.23	55.19
		Overall Error	37.90	37.50
	Test Set (331)	Yes (1)	17.71	31.25
		No (2)	68.35	57.55
		Overall Error	38.97	42.30
Combination of Web Usage and Demographic Variables (k=17 for normalized data & k=14 for non-normalized data)	Training Set (828)	Yes (1)	12.60	13.02
		No (2)	48.84	63.37
		Overall Error	27.66	33.94
	Validation Set (496)	Yes (1)	14.44	14.44
		No (2)	51.42	67.92
		Overall Error	30.24	37.30
	Test Set (331)	Yes (1)	14.58	22.40
		No (2)	58.27	68.35
		Overall Error	32.93	41.69

Appendix C. Error Rates for Artificial Neural Network Technique

Artificial Neural Network (Hidden Layer = 2)		Class Label	Standard Partitioning	
			With Data Normalization	Without Data Normalization
			% Error	% Error
Web Usage Variables	Training Set (828)	Yes (1)	34.92	27.27
		No (2)	15.70	29.65
		Overall Error	26.93	28.26
	Validation Set (496)	Yes (1)	40.49	30.63
		No (2)	23.11	34.91
		Overall Error	33.06	32.46
	Test Set (331)	Yes (1)	36.98	28.65
		No (2)	23.74	36.69
		Overall Error	31.42	32.02
Demographic Variables	Training Set (828)	Yes (1)	45.25	100.00
		No (2)	21.51	0.00
		Overall Error	35.39	58.45
	Validation Set (496)	Yes (1)	49.65	100.00
		No (2)	27.83	0.00
		Overall Error	40.32	57.26
	Test Set (331)	Yes (1)	50.52	100.00
		No (2)	28.78	0.00
		Overall Error	41.39	58.01
Combination of Web Usage and Demographic Variables	Training Set (828)	Yes (1)	33.47	100.00
		No (2)	13.37	0.00
		Overall Error	25.12	58.45
	Validation Set (496)	Yes (1)	37.68	100.00
		No (2)	25.00	0.00
		Overall Error	32.26	57.26
	Test Set (331)	Yes (1)	39.58	100.00
		No (2)	23.02	0.00
		Overall Error	32.63	58.01