International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

# Supervised Learning Algorithms of Machine Learning: Prediction of Brand Loyalty

**Nagaraju Kolla, M. Giridhar Kumar**

***ABSTRACT:*** ***The present research explores the loyalty prediction problem of a brand through supervised learning algorithms of classifications: logistic regression, decision tree, support vector machine, bayes algorithm and K-nearest neighbors (KNN) algorithm. 265 customers' FMCG loyalty sample data were taken and variables of the data set include; loyalty status, gender, family size, age, frequency of purchase, and FMCG purchase. Data have been analyzed with the help of Python packages such as Pandas (Data analysis), Numpy (Numerical calculation), Matplotlib (Visualization), and Sklearn (Modeling). Among the supervised classification algorithms, logistic regression has outperformed than other techniques.***

***KEYWORDS:*** ***Brand loyalty, FMCG, Logistic regression, Decision tree, Support vector machine, Bayes algorithm and K-nearest neighbors (KNN).***

## I.INTRODUCTION

**BRAND LOYALTY:** Brand loyalty reduces marketing costs (Levins, 2009), enhances the revenue (Reichheld & Sasser, 1990), creates positive perceptions (Gregg & Walczak, 2010; Kolla, Nagaraju, Geetha V.G. Devi, and R. Varaprasad, 2017), and finally contributes to returns of investment (Mizik & Jacobson, 2006). Jacoby and Chestnut (1978) defines brand loyalty as a biased decision making over alternative brands. Brand loyalty recently has evolved from single facet to composite.

*One dimensional Brand Loyalty:* The behavioural approach of the brand loyalty treats loyalty as a behavior (Kahn et al., 1986). According to behavioural approach, the customer who purchases the same brand over period of time can be treated as loyal. Behavioural loyalty is a stochastic process hence it is difficult to understand (Rundle-Thiele, 2005). But purchase behaviour is not the only criteria to define the loyalty because customer may purchase the same product for different reasons.

*Two Dimensional Loyalty:* For eliminating the drawback of behavioural loyalty, Day (1969) has introduced the two dimensional loyalty: attitudinal and behavioural. Attitudinal loyalty represents the customer beliefs, attitudes, and opinions to a selected brand (Pritchard et al., 1992). But attitudinal loyalty raised the criticism because it lacks the predicting power of actual purchase (Bennett & Rundle-Thiele, 2002).

**Manuscript published on 30 September 2019.**
*****Correspondence Author(s)
**Dr. Nagaraju Kolla,** Assistant Professor, G. Pullaiah College of Engineering & Technology (Autonomous) Kurnool/ Andhra Pradesh/ India
**Dr. M. Giridhar Kumar,** Professor and Head of the Department, Department of Management Studies, G. Pullaiah College of Engineering and Technology (Autonomous), Kurnool, A.P. India



*Three Dimensional Loyalty:* Oliver (2015) defined three dimensional nature of brand loyalty by including: cognitive, affective/attitude, and behavioral intentions. Cognitive loyalty is a new dimension and it talks about switching costs, the evaluation of the brand attributes and brand value.

*Four dimensional Loyalty:* Finally, brand loyalty has reached to four dimensions and it includes: cognitive, affective (attitudes), conative (deeply held commitment or repeat purchase) and action (behaviour). It is rare to purchase only one brand, possible to buy a brand again and again but no commitment to it. Hence conative eliminates the confusion over brand loyalty and other repeat purchase constructs.

**SUPERVISED LEARNING ALGORITHMS**: Machine learning is a recent advancement in data science and it can solve the complex problems of business. Computers/systems in machine learning can learns from the past data, and practice. Machine learning algorithms are of two types: supervised (regression or classification) and unsupervised learning algorithms (regression or classification). Supervised learning algorithm is a vital research topic in machine learning (Sun and Reddy 2013; Wu et al. 2014; Zhu et al. 2016). Recently, a great interest was devoted especially for the classification (Luo et al. 2016; Zhu et al. 2016). The present research explores the loyalty prediction problem of a brand through supervised learning algorithms (classification) such as: logistic regression, decision tree, support vector machine, bayes algorithm and K-nearest neighbors (KNN) algorithm.

*Regression analysis:* Regression can estimate the association between one response and a set of independent variables. Regression includes many models such as: Linear Regression (Dependent & independent variable-continuous), Logistic Regression (Dependent variable-binary (0/ 1)), Polynomial Regression (Power of independent variable is more than 1), Ridge Regression (ridge regression shrinks the coefficients and it helps to reduce the model complexity (L2 penalty)), Lasso Regression (Deals the multi-co linearity and also helps in features selection-feature engineering, (L1 penalty)), and Elastic Net Regression( Both L1(set some coefficients to zero) and L2 (shrink coefficients) penalty) . The present research has considered logistic regression for brand loyalty prediction. Logistic regression is a most cited statistical tool and can be used for forecasting the conditional probabilities (Berkson, 1944).

*Decision tree:* Decision tree have the applications in both supervised (regression) and non supervised (classification) learning. Which means it can support both categorical and continuous data.

Decision tree visualize the decisions with tree like model and the normal tree structure consists of three elements: internal node (variable/ feature), branches (output/result) and leaf node (labels of class). The present research uses the classification variant of decision tree.

*Support vector machine:* Support Vector Machine is a supervised machine learning algorithm and it has the applications in classification, predictions, face recognitions etc. (Fernandez, 1999; Tefas et al., 2001). SVM statistically minimizes the structural risk and help for better classification and predictions. SVM perform the classification by using the hyper planes. The error divides the SVM into four types: C-SVM classification, nu-SVM classification, epsilon-SVM regression, nu-SVM regression. C and nu SVM are equally plays vital role in classification but in terms of regularization nu is easier than C. Scikit-learn was a widely accepted library which can implement SVM algorithm. The non-linearly separable data, uses Kernel SVM which is a modified version of SVM.

*Bayes algorithm:* Bayes Algorithm is a classification technique runs through Bayes' Theorem. Distributions values of each class divide the Naïve Bayes Classifier into; the Gaussian Naïve Bayes (features-continuous), Multinomial Naïve Bayes (features-multinomial variables) and Bernoulli Naïve Bayes (features-binary variables). The present research uses Bernoulli Naïve Bayes.

*The K-nearest neighbors (KNN) algorithm:* The K-nearest neighbors (KNN) are a non- parametric supervised machine learning algorithm which classifies the data but doesn't work with high dimensional data. It is non-parametric, hence doesn't require underlying assumptions about data distribution. It is also called as non parametric lazy learning algorithm.

## II.METHODOLOGY

The present research explores the loyalty prediction problem of the brand through supervised learning algorithms of classification such as: logistic regression, decision tree, support vector machine, bayes algorithm and K-nearest neighbors (KNN) algorithm. For analyzing the supervised learning algorithms of classification the following methodology have applied (Figure-1).

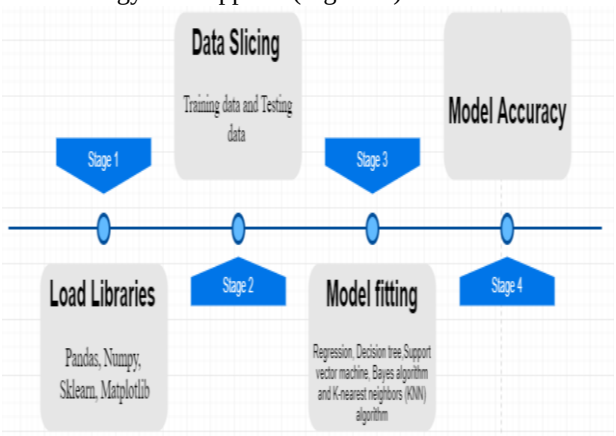


**Figure-1: Methodology of Proposed Research**

A sample of 265 FMCG loyal customers has taken from Kurnool/ Andhra Pradesh/ India. The data set of the research consists of; loyalty status, gender, family size, age, frequency of purchase, and FMCG purchase.

## DATA ANALYSIS AND INTERPRETATION

Supervised learning algorithms have been executed through three step processes loading of packages, data slicing and accuracy calculation. The first two steps are common for all the supervised techniques and last step varies along with the chosen model (https://dataaspirant.com accessed the site on 20/07/2019).

**I. Loading packages**

Python packages of Pandas (Data analysis), Numpy (Numerical calculation), Matplotlib (Visualization), and Sklearn (Modeling) are required for execution of supervised learning algorithms. Hence with the help of following code the packages were imported.

```
# Loading of required packages
pandas
numpy
matplotlib.pyplot
```

**II. Data Splitting:** Once required packages were loaded into the python the data set has to split into train data set and test dataset by means of scikit-learn (sk-learn) method. The research uses 80% of the research data as training data and rest of the 20% as testing data.

```
# Data splitting
# The research splits the data set with the help of the following code
dataset = pd.read_csv(fmcg_loyalty)
training_features = ['age', 'gender', ' family size ', ' frequency of purchase', ' FMCG purchase']
target = ' loyalty status '
```

## III.III. ACCURACY CALCULATION

The final step of the supervised learning algorithm execution is accuracy calculation.

*Logistic regression:* The accuracy level of logistic regression is presented below;

```
# Accuracy of logistic regression model

print "Test Accuracy :: ", test_accuracy
```

```
Test Accuracy: 0.94716981
```

*Decision tree:* Murthy (1998) have explained the significance of decision trees in the field of machine learning. Decision Tree Classifier (DTC) is the classifier function and it includes: criterion (indicates the excellence of a split: gini value), splitter (how to choose split at node-best value), max features (no. of features), max depth(depth of the tree), min samples split ( 2 for the present research), min samples leaf (1 value), max leaf node( none), and min impurity split(threshold).

*Retrieval Number: J94980881019/19©BEIESP*
*DOI: 10.35940/ijitee.J9498.0981119*
*Journal Website: www.ijitee.org*



*Published By:*
*Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)*

```
# Accuracy of Decision tree

print "Accuracy is ",
accuracy_score(y_test,y_pred)*….
```

```
Accuracy = 0.90566038
```

Decision tress predicts the models with precise level of accuracy (Hang Yang, Fong, 2011). But the accuracy of the decision tree acts under the control of tree complexity (Breiman et al, 2007, Hancock et al., 1996, Hyafil and Rivest, 1976).

*Support vector machine (SVM):* Vapnik (1995) have first introduced the term Support Vector Machine (SVM) in machine learning research. Since then it attracted the high level of interest in machine learning (V. Vapnik, 1995). Recent studies also claimed the accuracy of SVM in classification (Drucker N, Donghui W, Vapnik VN, 1999; Hua S, Sun Z , 2001; Strauss DJ, Steidl G, 2002) .

```
# Accuracy of Support vector machine (SVM)
SVC= 0.93962264
```

*Bayes algorithm:* In the literature, the Bayes algorithm have applied in text classification ( G. Feng, J. Guo, B.-Y. Jing, & L. Hao, 2012), image processing ( Z. Zhang, Q. Zhu, & Y. Xie, 2012, M. C. Yang, C. S. Huang, J. H. Chen, & R. F. Chang, 2012 ), and fault prediction (C. Catal & B. Diri, 2009).

```
# Accuracy of Bayes algorithm
clf = Bayes()
```

```
Accuracy = 0.90188679
```

*The K-nearest neighbors (KNN) algorithm:* It is a non-parametric supervised machine learning algorithm and the accuracy of model is executed with the following code. The result is furnished in the below;

```
# Accuracy of the K-nearest neighbors
(KNN) algorithm
# With respect to classification
algorithms score method provides the
accuracy.

knn.score(X_test,y_test)
```

```
Accuracy= 0.91320755
```

## IV.CONCLUSION

The brand loyalty prediction of FMCG with different machine learning algorithms like regression, decision tree, support vector machine, bayes algorithm and K-nearest neighbors (KNN) algorithm are executed for the 265 sample dataset.

**Table-1: Output on Brand Loyalty Prediction**

| Method | Classified | Misclassified | Accuracy (%) |
|---|---|---|---|
| Decision tree | 240 | 25 | 0.90566038 |
| K-NN | 242 | 23 | 0.91320755 |
| Logistic regression | 251 | 14 | 0.94716981 |
| Naïve Bayes | 239 | 26 | 0.90188679 |
| SVM | 249 | 16 | 0.93962264 |

**Source:** Primary research

Among the supervised classification algorithms, logistic regression has outperformed than other techniques. The research contributes for the better understanding of FMCG brand loyalty and is the ultimate dream of any branding professional.

## REFERENCES

1. Anbarasi, M., Anupriya, E., & Iyengar, N. (2010). Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm. International Journal of Engineering Science and Technology .
2. Bennett, R., & Rundle-Thiele, S. (2002). A comparison of attitudinal loyalty measurement approaches. Journal of Brand Management .
3. Berkson, J. (1944). Application of the Logistic Function to Bio-Assay. Journal of the American Statistical Association .
4. Breiman, L., Friedman, J., Olshen, R., & Stone, C. (2017). Classification and regression trees.
5. Catal, C., & Diri, B. (2009). Investigating the effect of dataset size, metrics sets, and feature selection techniques on software fault prediction problem. Information Sciences .
6. Chapelle, O., Haffner, P., & Vapnik, V. (1999). Support vector machines for histogram-based image classification. IEEE Transactions on Neural Networks .
7. Day, G. (1976). A Two-Dimensional Concept of Brand Loyalty. In G. Day.
8. Drucker, H., Wu, D., & Vapnik, V. (1999). Support vector machines for spam categorization. IEEE Transactions on Neural Networks .
9. Feng, G., Guo, J., Jing, B., & Hao, L. (2012). A Bayesian feature selection paradigm for text classification. Information Processing and Management .
10. Fernandez, R. (1999). Predicting time series with a local support vector regression machine. Proceedings of the ECCAI Advanced Course on … .
11. Gregg, D., & Walczak, S. (2010). The relationship between website quality, trust and price premiums at online auctions. Electronic Commerce Research .
12. Hancock, T., Jiang, T., Li, M., & Tromp, J. (1996). Lower Bounds on Learning Decision Lists and Trees. Information and Computation .
13. Hua, S., & Sun, Z. (2001). A novel method of protein secondary structure prediction with high segment overlap measure: Support vector machine approach. Journal of Molecular Biology .
14. Hyafil, L., & Rivest, R. (1976). Constructing optimal binary decision trees is NP-complete. Information Processing Letters .
15. Kahn, B., Kalwani, M., & Morrison, D. (2006). Measuring Variety-Seeking and Reinforcement Behaviors Using Panel Data. Journal of Marketing Research .

16. Kolla, N., Devi, G., & Varaprasad, R. (2017). Are certain brand personality dimensions more valuable at driving loyalty than others? International Journal of Applied Business and Economic Research .
17. Luo, C., Zeng, J., Yuan, M., Dai, W., & Yang, Q. (2016). Telco User Activity Level Prediction with Massive Mobile Broadband Data. ACM Transactions on Intelligent Systems and Technology .
18. Oliver, R. (2015). Satisfaction: A Behavioral Perspective on the Consumer.
19. Pritchard, M., Howard, D., & Havitz, M. (1992). Loyalty measurement: A critical examination and theoretical extension. Leisure Sciences .
20. Reichheld, F., & Sasser, W. (1990). Zero defections: quality comes to services. Harvard business review .
21. Rundle-Thiele, S. (2005). Exploring loyal qualities: Assessing survey-based loyalty measures. Journal of Services Marketing .
22. Sain, S., & Vapnik, V. (2006). The Nature of Statistical Learning Theory. Technometrics .
23. Strauss, D., & Steidl, G. (2002). Hybrid wavelet-support vector classification of waveforms. Journal of Computational and Applied Mathematics .
24. Sun, J., & Reddy, C. (2013). Big data analytics for healthcare. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13.
25. Tefas, A., Kotropoulos, C., & Pitas, L. (2001). Using Support Vector Machines to enhance the performance of elastic graph matching for frontal face authentication. IEEE Transactions on Pattern Analysis and Machine Intelligence .
26. Webb, G., Boughton, J., & Wang, Z. (2005). Not so naive Bayes: Aggregating one-dependence estimators. Machine Learning .
27. Xindong Wu, Gong-Qing Wu, W. (2014). Data Mining with Big Data Xindong. Revista Argentina de Dermatologia .
28. Yang, H., & Fong, S. (2011). Optimized very fast decision tree with balanced classification accuracy and compact tree size. Proceedings - 3rd International Conference on Data Mining and Intelligent Information Technology Applications, ICMIA 2011.
29. Yang, M., Huang, C., Chen, J., & Chang, R. (2012). Whole Breast Lesion Detection Using Naive Bayes Classifier for Portable Ultrasound. Ultrasound in Medicine and Biology .
30. Zhang, Z., Zhu, Q., & Xie, Y. (2012). A novel image matting approach based on naive Bayes classifier. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).
31. Zhu, X., Li, X., Zhang, S., Ju, C., & Wu, X. (2017). Robust Joint Graph Sparse Coding for Unsupervised Spectral Feature Selection. IEEE Transactions on Neural Networks and Learning Systems .