

Composite Model Fabrication of Classification with Transformed Target Regressor for Customer Segmentation using Machine Learning

Rincy Merlin Mathew, R. Suguna, M. Shyamala Devi



Abstract: In Current internet world, the customers prefer to buy the products through online rather than spending their time on show rooms. The online customers of wine increases day by day due to the availability of high brands in online sellers. So the customers buy the wine products based on the product description and the satisfaction of other customers those who have bought before. This makes the industries to focus on machine learning that concentrates on target transformation of the dependent variable. This paper endeavor to forecast the customer segmentation for the wine data set extracted from UCI Machine learning repository. The raw wine data set is subjected to target transformation for various classifiers like Huber Regressor, SGD Regressor, RidgeCV Regression, Logistic RegressionCV and Passive Aggressive Regressor. The performance of the various classifiers is analyzed with and without target transformation using the metrics like Mean Absolute Error and R2 Score. The implementation is done in Anaconda Navigator with Python. Experimental results shows that after applying target transformation RidgeCV Regression is found to be effective with the R2 Score of 82% and Mean Absolute Error of 0.0 compared to other classifiers.

Index Terms: Machine Learning, Target Transformation, Classifier, Mean Absolute Error and R2 Score.

I. INTRODUCTION

Customer behavior must be predicted in advance for the growth of the profit and the total income of the organization. The profit of any company has to be increased year by year by projecting the model of target transformation. The machine learning models are also applied to concentrate only on the target transformation of the dependent variable so as to forecast the current needs and the future needs of the customer buying the product. The paper is organized in such a way that Section 2 deals with the related works. Section 3 discuss about the proposed work followed by the implementation and Performance Analysis in Section 4. The paper is concluded with Section 5.

II. RELATED WORK

A. Literature Review

The customer relationship management is greatly needed for any business to survive in the current market world. The utilization charge of wine was evaluated using various factors like such as manufactured goods involvement, biased awareness, delicate qualities and socio demography [1].

Due to the growth in the online shopping, the customers wish to buy the high quality wine through online web portal shopping. In this scenario, the customers just view the quality of the wine only through the ingredients present in the wine [2]. The various wine brands have worth in their improvement and the current market is highly competitive [3]. The survey is attempted to design by analyzing the various wine data set attributes and the behavior patterns of the wine date set. This survey is used in predicting the customer's behavior so as to increase the profit of the organization. This is attempted with the motive of finding and analyzing the customer's intentions in buying the product. [4] The ingredients in the wine product also greatly influence the people to buy the product and they attempted to find the essential ingredients by analyzing the feedback and the past history of product sales [5]. The composition of the mixing of the product also greatly influences the people in buying the product. They attempt to find the combination of the ingredients in forming the wine product and shown that the sales of the wine product got increased due to exact combination [6]. The model was created to develop each wine product and the standardization of the product is done based on the benchmark survey level of the wine consumers and the prediction is done with the benchmark data [7]. The application of Feature Selection and Feature Extraction on wine data set is done to predict the target [8]-[12].

III. PROPOSED WORK

In our proposed work, the wine data set is applied to foresee the customer segmentation based on the target transformation. Our implementation in this paper is folded in six ways.

- (i) Firstly, the correlation matrix is created that depicts the relativity of each of the attributes present in the wine data set.
- (ii) Secondly, the analysis of Customer segment based on the target distribution.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

Rincy Merlin Mathew*, Lecturer, Department of Computer Science, College of Science and Arts, Khamis Mushayt, King Khalid university, Abha, Asir, Saudi Arabia.

R. Suguna, Professor, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

M. Shyamala Devi, Associate Professor, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Composite Model Fabrication of Classification with Transformed Target Regressor for Customer Segmentation using Machine Learning

- (iii) Thirdly, the wine dataset is applied to target transformation for various classifiers like Huber Regressor, SGD Regressor, RidgeCV Regression, Logistic RegressionCV and Passive Aggressive Regressor
- (iv) Fourth, the performance of the various classifiers is analyzed with and without target transformation using the metrics like Mean Absolute Error and R2 Score.

- 7. Flavanoids
- 8. Nonflavanoid phenols
- 9. Proanthocyanins
- 10. Color intensity
- 11. Hue
- 12. OD280/OD315 of diluted wines
- 13. Proline Churn
- 14. Customer Segment - Dependent Attribute

A. System Architecture

The propose architecture of our work is shown in Fig. 1

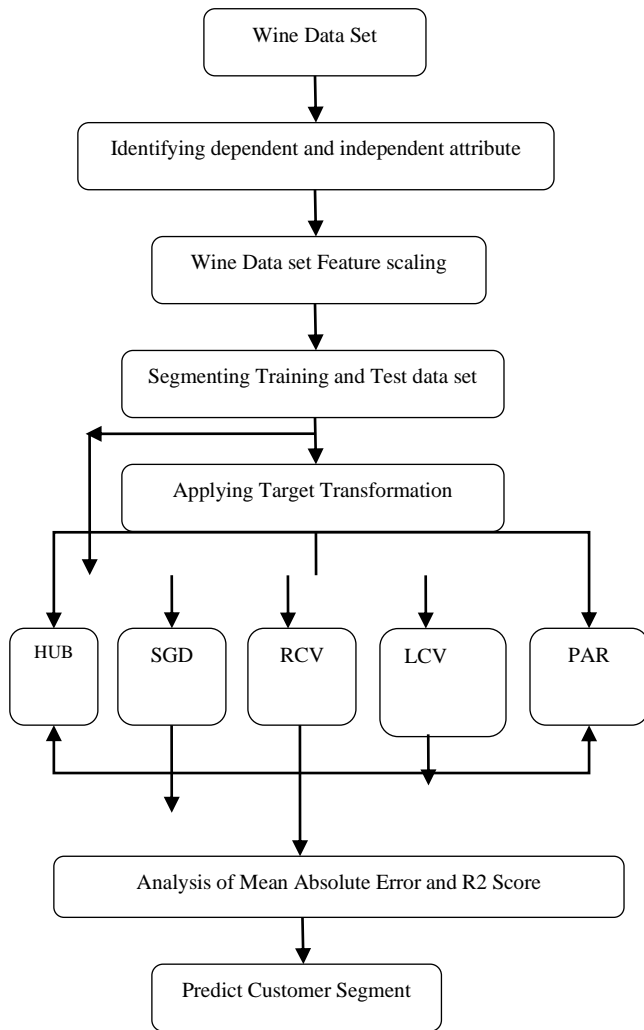


Fig. 1 System Architecture

IV. IMPLEMENTATION AND PERFORMANCE ANALYSIS

A. Customer Segmentation Prediction

The Wine dataset from UCL ML Repository is used for implementation with 13 independent attribute and 1 Customer Segment dependent attribute. The attribute are shown below.

1. Alcohol
2. Malic acid
3. Ash
4. Alcalinity of ash
5. Magnesium
6. Total phenols

The relativity between the attributes of the wine data set is depicted as a correlation matrix and is shown in Fig. 2.

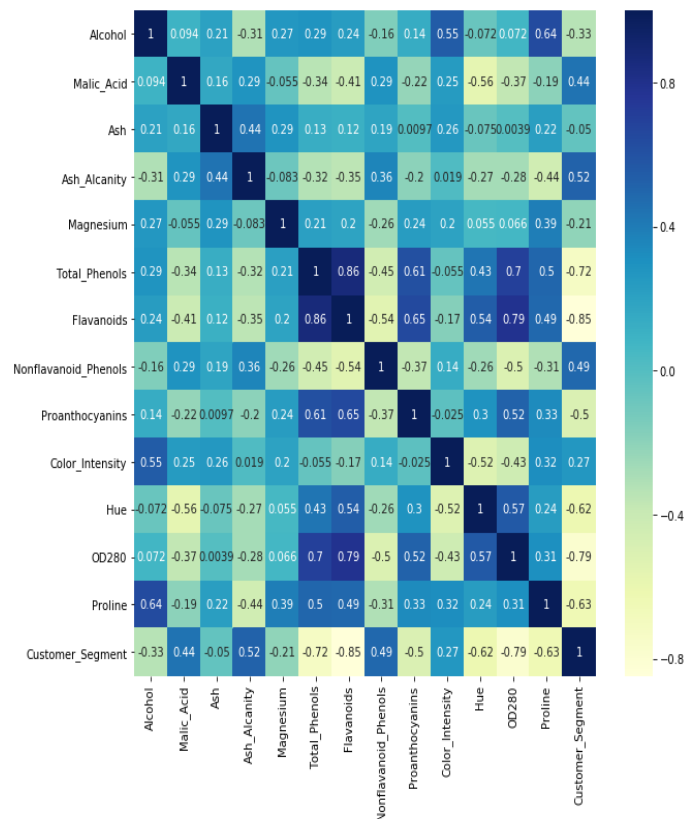


Fig. 2 Correlation Matrix of Wine data set

The raw wine data set is projected with the distribution of the customer segment target variable and is shown in Fig. 3.

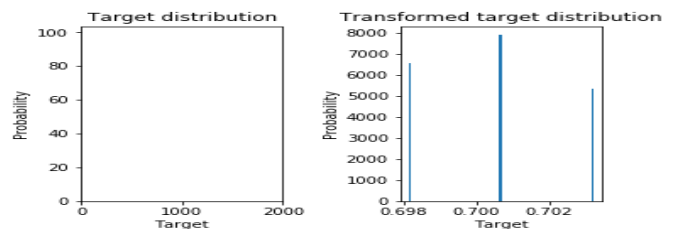


Fig. 3. Customer Group Segmentation with Target Distribution

The distribution of customer segment target variable of the wine data set is done for the Huber Regressor with and without target transformation and is shown in Fig. 4 and the performance metric is also analyzed for Mean Absolute Error and R2 Score.

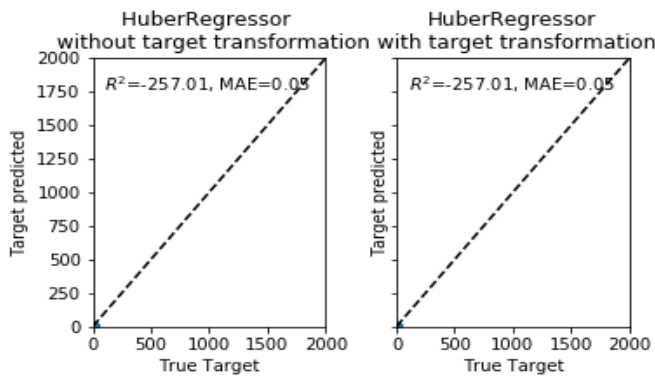


Fig. 4. Huber Regressor with and without Target Transformation

The distribution of customer segment target variable of the wine data set is done for the SGD Regressor with and without target transformation and is shown in Fig. 5 and the performance metric is also analyzed for Mean Absolute Error and R2 Score.

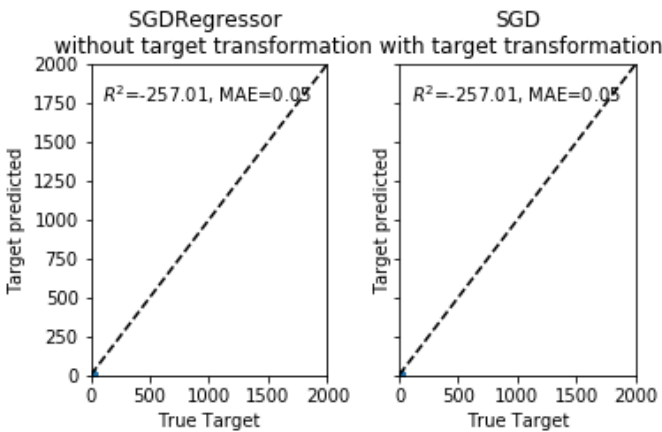


Fig. 5. SGD Regressor with and without Target TRANSFORMATION

The distribution of customer segment target variable of the wine data set is done for the RidgeCV Regressor with and without target transformation and is shown in Fig. 6 and the performance metric is also analyzed for Mean Absolute Error and R2 Score.

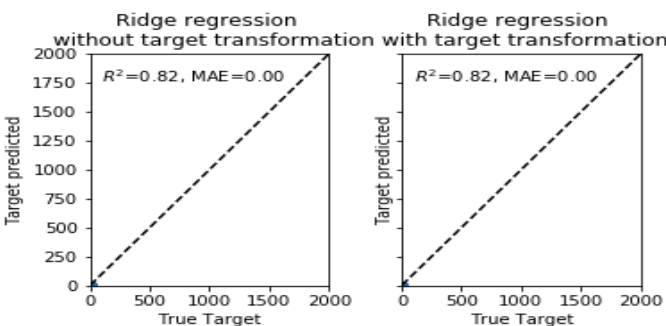


Fig. 6. Ridge Regressor with and without Target Transformation

The distribution of customer segment target variable of the wine data set is done for the Logistic RegressionCV classifier with and without target transformation and is shown in Fig.7 and the performance metric is also analyzed for Mean Absolute Error and R2 Score.

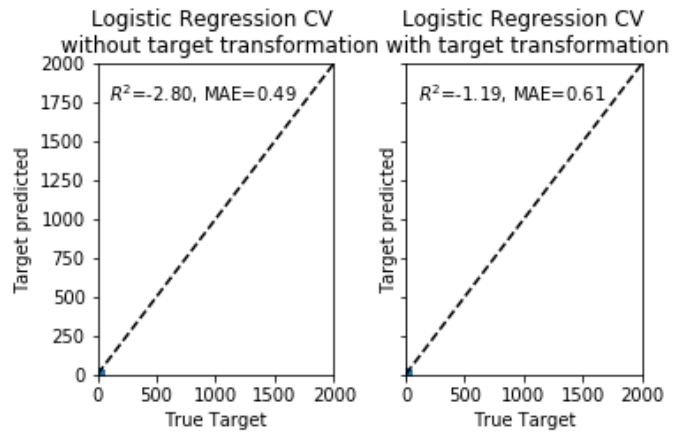


Fig. 7. Logistic Regressor CV with and without Target Transformation

The distribution of customer segment target variable of the wine data set is done for the Passive Aggressive Regressor classifier with and without target transformation and is shown in Fig. 8 and the performance metric is also analyzed for Mean Absolute Error and R2 Score.

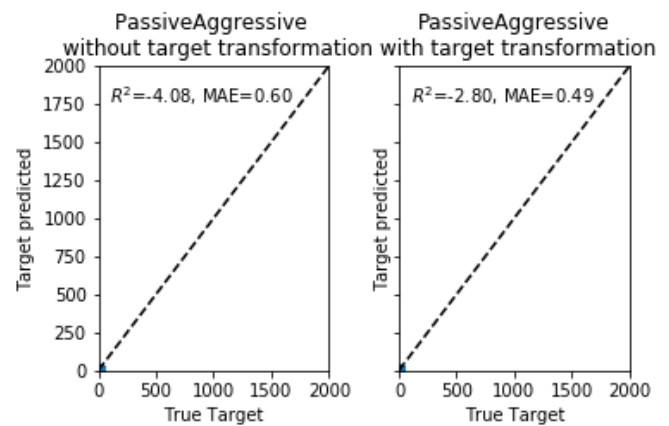


Fig. 8 Passive Aggressive Regressor with and without Target Transformation

The the wine dataset is applied to target transformation for various classifiers like Huber Regressor, SGD Regressor, RidgeCV Regression, Logistic RegressionCV and Passive Aggressive Regressor The performance of the various classifiers is analyzed with and without target transformation using the metrics like Mean Absolute Error and R2 Score and is shown in the Table.1 and Table. 2.

Table. 1 Performance Comparison of MAE and R2 Score without target transformation

Regressor Models	Without fitting Transformed Target Regressor	
	MAE	R2 Score
Huber Regressor	0.05	-257.01
SGD Regressor	0.05	-257.01

Composite Model Fabrication of Classification with Transformed Target Regressor for Customer Segmentation using Machine Learning

RidgeCV Regressor	0.00	0.82
Logistic Regressor	0.49	2.80
Passive Aggressive Regressor	0.60	4.08

Table. 2 Performance Comparison of Comparison of MAE and R2 Score after target transformation

Regressor Models	Fitting with Transformed Target Regressor	
	MAE	R2 Score
Huber Regressor	0.05	-257.01
SGD Regressor	0.05	-257.01
RidgeCV Regressor	0.00	0.82
Logistic Regressor	0.61	1.19
Passive Aggressive Regressor	0.49	2.80

The Performance analysis of the metrics like Absolute Error and R2 Score without target transformation for various classifiers is shown in Fig. 9-Fig. 10.

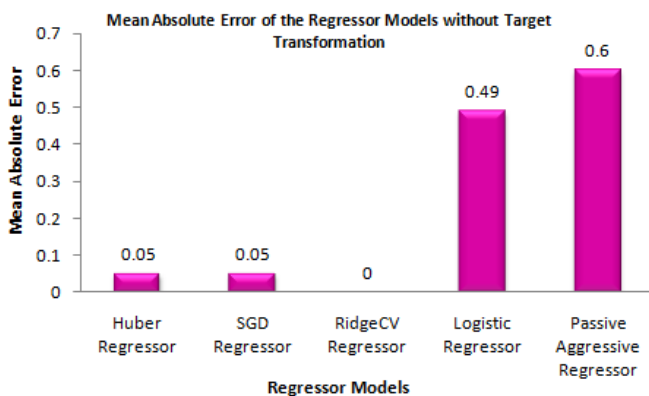


Fig. 9 Mean Absolute Error of Regressor models without Target Transformation

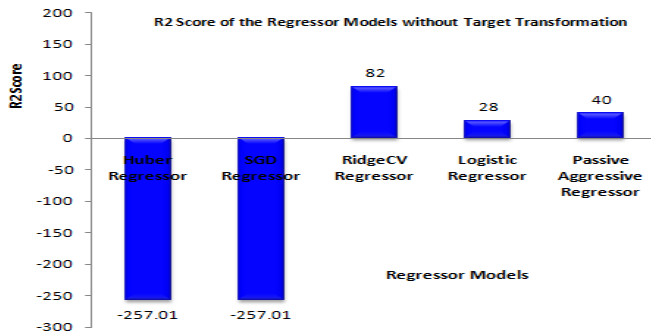


Fig. 10 R2 Score of Regressor models without Target Transformation

The Performance analysis of the metrics like Absolute Error and R2 Score without target transformation for various classifiers is shown in Fig. 11-Fig. 12.

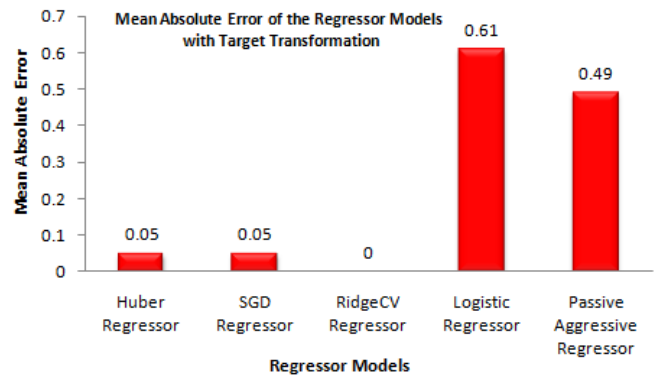


Fig. 11 Mean Absolute Error of Regressor models with Target Transformation

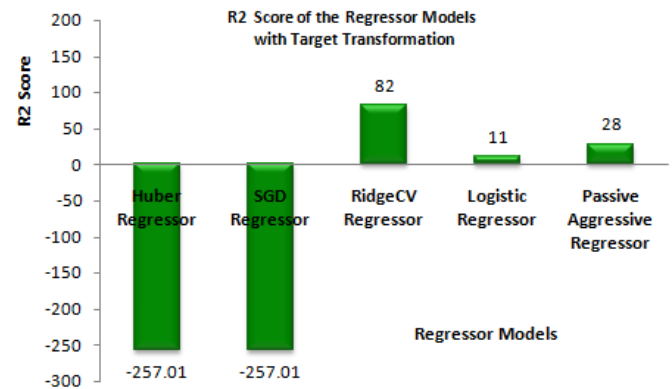


Fig. 12 R2 Score of Regressor models with Target Transformation

V. CONCLUSION

This paper attempts to predict the customer segmentation of the wine data set. The raw wine data set is subjected to target transformation for various classifiers like Huber Regressor, SGD Regressor, RidgeCV Regression, Logistic RegressionCV and Passive Aggressive Regressor. The performance of the various classifiers is analyzed with and without target transformation using the metrics like Mean Absolute Error and R2 Score. Experimental results shows that after applying target transformation RidgeCV Regression is found to be effective with the R2 Score of 82% and Mean Absolute Error of 0.0 compared to other classifiers.

REFERENCES

1. David Cox, "Predicting Consumption, Wine Involvement and Perceived Quality of Australian Red Wine", *Journal of Wine Research.*, vol. 20, no. 3, 2009, pp. 209-229.
2. Constanza Bianchi, "Consumer Brand Loyalty in the Chilean Wine Industry", *Journal of Food Products Marketing.*, vol. 21, no. 4, 2015, pp. 442-460.
3. D. Veena Parboteeah, D. Christopher Taylor, and A. Nelson Barber, "Exploring impulse purchasing of wine in the online environment", *Journal of Wine Research.*, vol. 27, no. 4, 2016, pp. 322-339.
4. Hyojin Kim, and A. Mark Bonn, "The Moderating Effects of Overall and Organic Wine Knowledge on Consumer Behavioral Intention", *Scandinavian Journal of Hospitality and Tourism.*, vol. 15, no. 3, 2015 pp. 295-310.

5. Johan Bruwer, Nicole Burrows, Sylvia Chaumont, Elton Li, and Anthony Saliba, "Consumer involvement and associated behaviour in the UK high-end retail off-trade wine market", *The International Review of Retail, Distribution and Consumer Research.*, vol. 24, no. 2, 2014, pp. 145-165.
6. Johan Bruwer, Justin Cohen, and Kathleen Kelley, "Wine involvement interaction with dining group dynamics, group composition and consumption behavioural aspects in USA restaurants", *International Journal of Wine Business Research.*, vol. 3, no.1, 2019, pp.12-28.
7. Gabriele Scozzafava, Francesca Gerini, Andrea Dominici, Caterina Contini, and Leonardo Casini. "Reach for the stars: The impact on consumer preferences of introducing a new top-tier typology into a PDO wine", *Wine Economics and Policy.*, vol. 7, no. 2, 2018, pp. 140-152.
8. M. Shyamala Devi, Rincy Merlin Mathew, and R. Suguna,"Attribute Heaving Extraction and Performance Analysis for the Prophecy of Roof Fall Rate using Principal Component Analysis", *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no.8, June 2019, pp. 2319-2323.
9. R. Suguna, M. Shyamala Devi, and Rincy Merlin Mathew, " Customer Churn Predictive Analysis by Component Minimization using Machine Learning", *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no.8, June 2019, pp. 2329-2333.
10. Shyamala Devi Munisamy, and Suguna Ramadass Aparna Joshi, "Cultivar Prediction of Target Consumer Class using Feature Selection with Machine Learning Classification", Springer's book series "Learning and Analytics in Intelligent Systems, Springer, LAIS vol. 3, pp. 604-612, 2019.
11. Suguna Ramadass, and Shyamala Devi Munisamy, Praveen Kumar P, Naresh P, "Prediction of Customer Attrition using Feature Extraction Techniques and its Performance Assessment through dissimilar Classifiers", Springer's book series "Learning and Analytics in Intelligent Systems, Springer, , LAIS vol. 3, pp. 613-620, 2019.
12. R.Suguna, M. Shyamala Devi, Rupali Amit Bagate, and Aparna Shashikant Joshi, "Assessment of Feature Selection for Student Academic Performance through Machine Learning Classification", *Journal of Statistics and Management Systems*, Taylor Francis, vol.22, no. 4, 25 June 2019, pp. 729-739. DOI:10.1080/09720510.2019.1609729
13. M. Shyamala Devi, Rincy Merlin Mathew, R. Suguna, "Feature Snatching and Performance Analysis for Connoting the Admittance Likelihood of student using Principal Component Analysis", *International Journal of Recent Technology and Engineering*, Volume-8 Issue-2, 30 July 2019, pp. 4800-4807.
14. R. Suguna, M. Shyamala Devi, Rincy Merlin Mathew, "Customer Segment Prognostic System by Machine Learning using Principal Component and Linear Discriminant Analysis", *International Journal of Recent Technology and Engineering*, Volume-8 Issue-2, 30 July 2019, pp. 6198-6203.