International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Machine Learning Based Target Transformation using Regressor Classification for Heart Disease Envisage

Rincy Merlin Mathew, M. Shyamala Devi, Shermin Shamsudheen

Abstract*: In the modern scenario of technological growth, the life style of an individual varies with the economic status. The world population is prone towards chronic deadly diseases due to the variety of food habits. The usages of electronic equipments have raised the population to waste their quality time towards exercise. The lack of physical activity has symptoms towards bad quality of life. With this background information, this paper concentrates on predicting the type of heart disease by applying target transformation using various machine learning regression models. This paper uses the Heart disease data set extracted from UCI Machine Learning Repository. The anaconda Navigator IDE along with Spyder is used for implementing the Python code. Our contribution is folded is folded in three ways. First, the data segregation is done and it is preprocessed to extract the relationship and dependency of each parameters. Second, the dataset is subjected to process to identify the target distribution of classes in the dependent variable. Third, the dataset is fitted to the Ridge regressor, Huber regressor, SGD regressor and PerceptronCV regressor by applying with and without target transformation. Fourth, dataset is feature scaled and then fitted to the Ridge regressor, Huber regressor, SGD regressor and PerceptronCV regressor by applying with and without target transformation. Fifth, the performance analysis is done by analyzing the Mean Absolute Error and R2 Score. Experimental results show that, the Perceptron regressor CV has the effectiveness with the mean absolute error of 1.00 and R2 score of 0.04 for the heart disease prediction.

***Index Terms*: Machine Learning, Regressor, Target transformation, MAE and R2 Score.**

I. INTRODUCTION

The world population is now endured with the various deadly chronic diseases because of the lack of metabolic activities. The lifestyle of the individual drastically changes due to the advancement and usage of electronic devices. There is a lack of awareness among the people towards healthy nutrition food habits. This leads to lack of exercises and physical activity for any individual. According to the survey, the disease affected patients are mostly suffered from cardio vascular disease and they are prone to extra infectious complications too. The health care industry is towards predicting the chronic disease in early stages and to treat them in advance. This methodology of diagnosing the chronic diseases in advance may help the physicians for treating the patients in advance and to reduce the serious complications. The heart disease may result in severe and sudden cardiac arrest at any time. This may lead to dangerous adverse health effect to the patients resulting to death. So the health care industry is spending their outcome towards predicting the heart disease in advance. Since the symptoms for patients affected with heart disease are enormous, this result in challenging issue for the researchers to find the exact clinical parameter that is responsible for the heart disease.

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 target transformation for the multi target regression models are built for attaining the continuous multiple target outputs. It uses the multiple regression chains which leads to building the individual regression models that uses the previous prediction of the regression models. It also finds the multiple correlation chain for predicting the multiple target transformed output. It can also be subjected to the non parametric statistical tests [1].

The decision support system has to be more effective for dignosing the heart disease. The neural network application is used in predicting the heart disease with the target transformation of the dependent target class variable. They have verified and used the various types of neural system with all the clinical parameters [2].

The neural network is extended to form the modular neural network for predicting the heart disease. It is also possible to perform the manual analysis method to predicting the heart disease. The same can be extended to be programmatic which is done by the segmental neural network which results in the intelligent expert structural model for analyzing the various patterns of heart disease [3]. The machine learning feature selection and feature extraction methods can be used for the prediction of any factor in different application can be learnt through this article [4] – [28].

Manuscript published on January 30, 2020.
* Correspondence Author
Rincy Merlin Mathew, Lecturer, Department of Computer Science, College of Science and Arts, Khamis Mushayt, King Khalid university, Abha, Asir, Saudi Arabia.
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.
Shermin Shamsudheen, Lecturer, Department of Computer Science, College of Computer Science & Information Systems, Jazan University, Saudi Arabia.

III. PROPOSED WORK

In our proposed work, the customer churn bank modelling data set is subjected to predict the customer churn analysis. Our implementation in this paper is folded in six ways.

(i) First, the data segregation is done and it is preprocessed to extract the relationship and dependency of each parameters.
(ii) Second, the dataset is subjected to process to identify the target distribution of classes in the dependent variable.
(iii) Third, the dataset is fitted to the Ridge regressor, Huber regressor, SGD regressor and PerceptronCV regressor by applying with and without target transformation.
(iv) Fourth, dataset is feature scaled and then fitted to the Ridge regressor, Huber regressor, SGD regressor and PerceptronCV regressor by applying with and without target transformation.
(v) Fifth, the performance analysis is done by analyzing the Mean Absolute Error and R2 Score.

A. System Architecture

The overall block diagram of the proposed work is shown in Fig. 1.

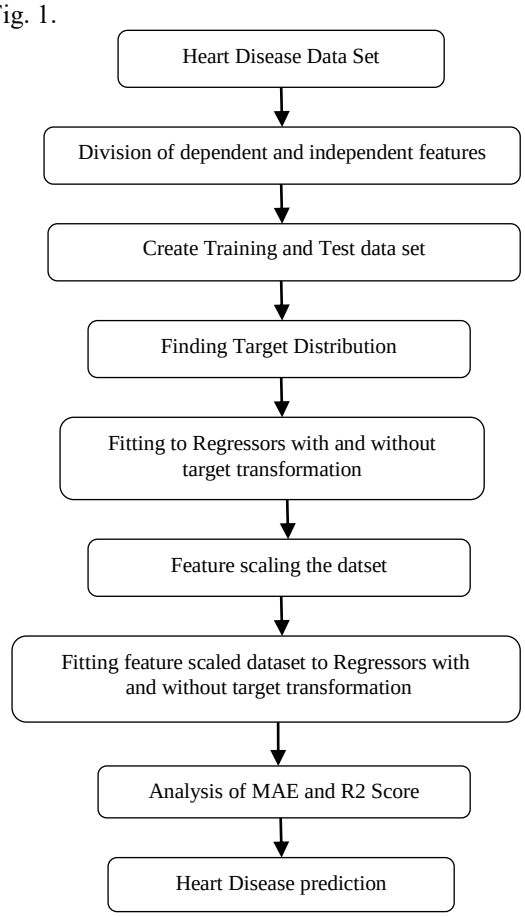

Fig. 1 Overall Design of Work

IV. IMPLEMENTATION AND PERFORMANCE ANALYSIS

A. Target Transformation Heart Disease Dataset

The Heart Disease dataset extracted from UCL ML Repository is used for implementation with 13 independent attribute and 1 diagnosis dependent attribute. The dataset consists of 779 individual's data. The attribute are shown below.

1.	Age
2.	Sex
3.	Chest-pain type
4.	Resting Blood Pressure
5.	Serum Cholestrol
6.	Fasting Blood Sugar
7.	Resting ECG
8.	Max heart rate
9.	Angina
10.	ST depression
11.	Peak exercise ST segment
12.	Number of major vessels
13.	Thal
14.	Diagnosis of heart disease - Dependent Attribute

The dependency of each parameters is shown in Fig. 2.

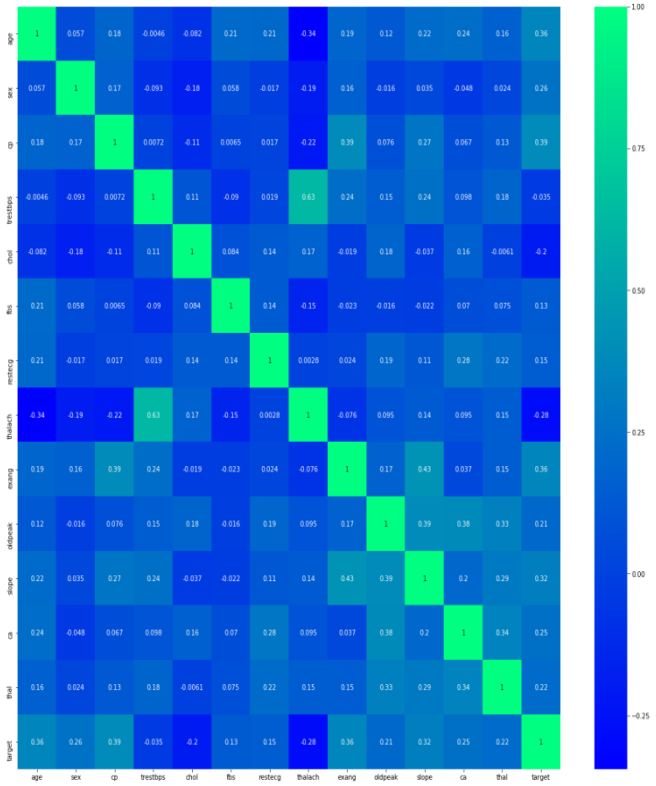

Fig. 2 Correlation Matrix of Heart Disease dataset

The classes of the dependent attribute is shown in Fig. 3.

Chest Pain Format	Desription
1	Typical Angina
2	Atypical Angina
3	Non - Anginal Pain
4	Asymptotic

Fig. 3. Pain Type Class Levels

The expansion of the dataset attributes shown in Fig. 4.

S.No	Attribute	Description
1.	Age	Age of the individual.
2.	Sex	Gender of the individual with the following format: 1 = male 0 = female.
3.	Chest-pain type	Type of chest-pain as in Table. 2.
4.	Resting Blood Pressure	Resting blood pressure value in mmHg (unit)
5.	Serum Cholestrol	serum cholestrol in mg/dl (unit)
6.	Fasting Blood Sugar	Fasting blood sugar value of an individual with 120mg/dl. If fasting blood sugar > 120mg/dl then : 1 (true) else : 0 (false)
7.	Resting ECG	0 = normal 1 = having ST-T wave abnormality 2 = left ventricular hypertrophy
8.	Max heart rate	Max Heart Rate
9.	Angina	1 = yes 0 = no
10.	ST depression	Value (integer or float).
11.	Peak exercise ST segment	1 = Upsloping 2 = Flat 3 = Downsloping
12.	Number of major vessels	Values (0-3) colored by flourosopy
13.	Thal	displays the Thalassemia : 3 = Normal 6 = Fixed Defect 7 = Reversable Defect
14.	Diagnosis of heart disease	0 = Absence 1,2,3,4 = Present.

Fig. 4. Heart Disease Dataset Design

The division of the target heart disease class type variable is shown in Fig.5.

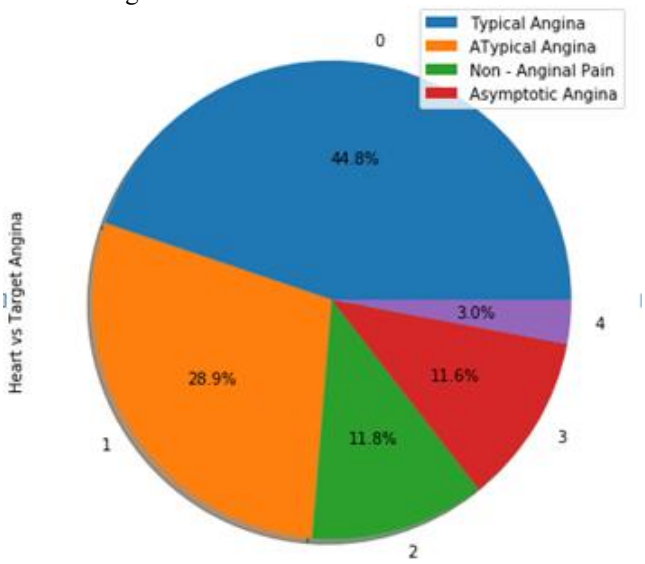

Fig. 5. Target Class Distribution

The dataset is fitted to the Ridge regressor, Huber regressor, SGD regressor and PerceptronCV regressor by applying with and without target transformation. The dataset is feature scaled and then fitted to the Ridge regressor, Huber regressor, SGD regressor and PerceptronCV regressor by applying with and without target transformation. is shown in the Fig. 6 – Fig 14.

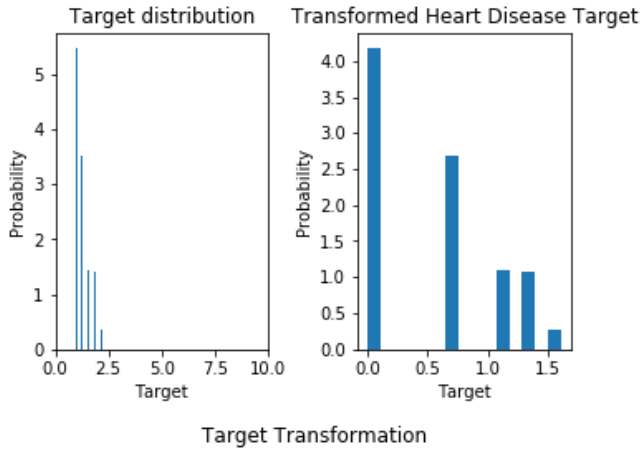

Fig. 6. Division of Target Class with Target transformation

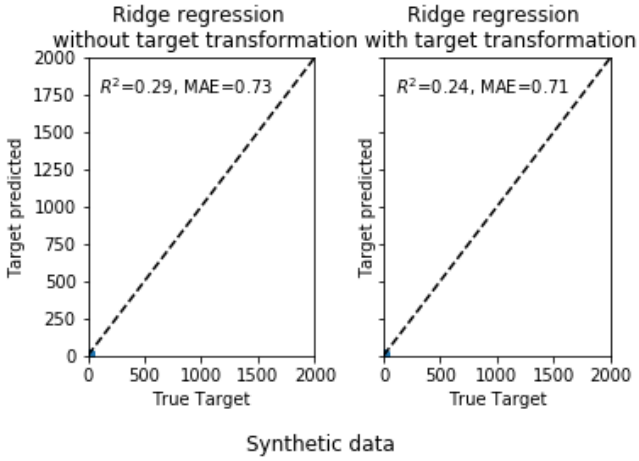

Fig. 7. Ridge Regression with Target transformation

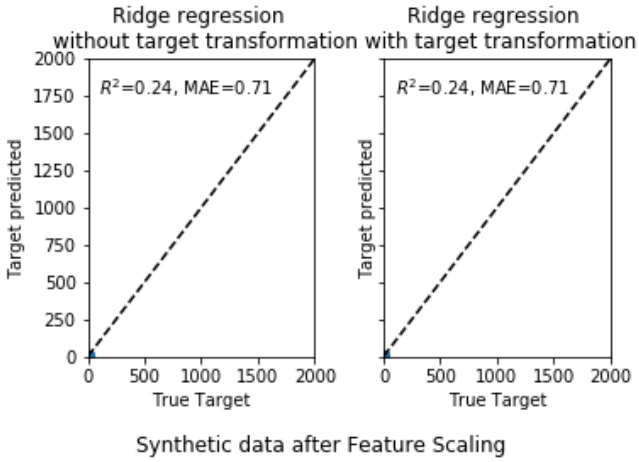

Fig. 8. Feature Scaled Ridge Regression with Target transformation

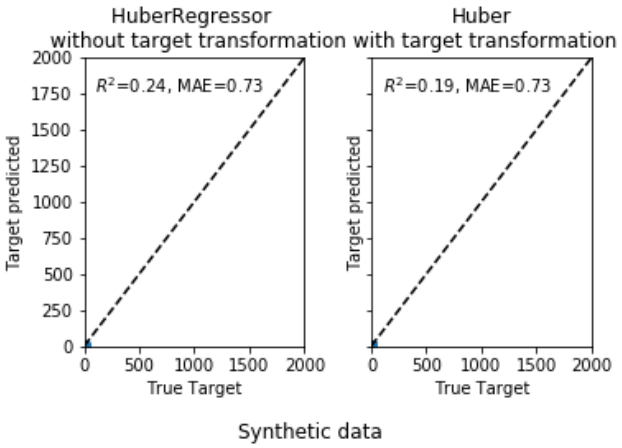

Fig. 9. Huber Regressor with Target transformation

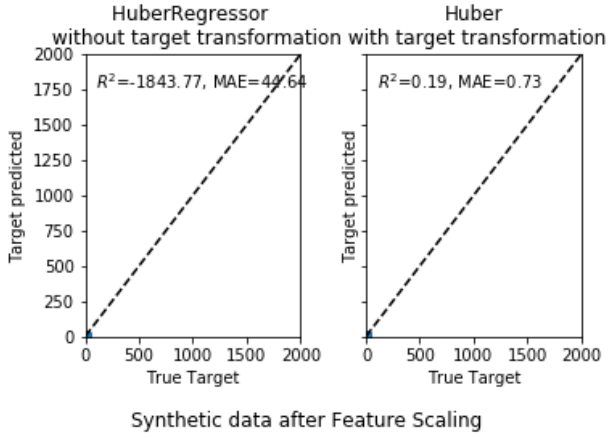

Fig. 10. Feature Scaled Huber Regressor with Target transformation

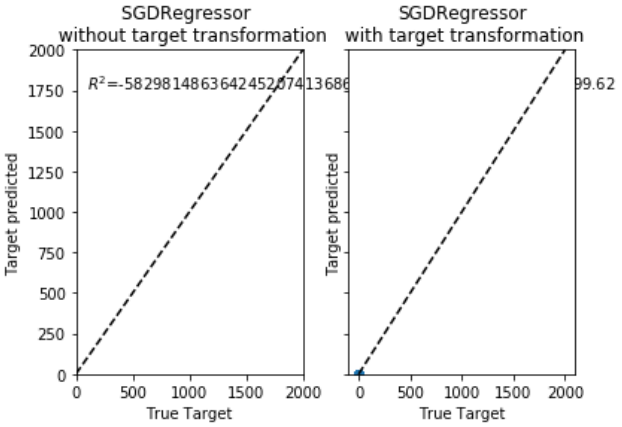

Fig. 11. SGD Regressor with Target transformation

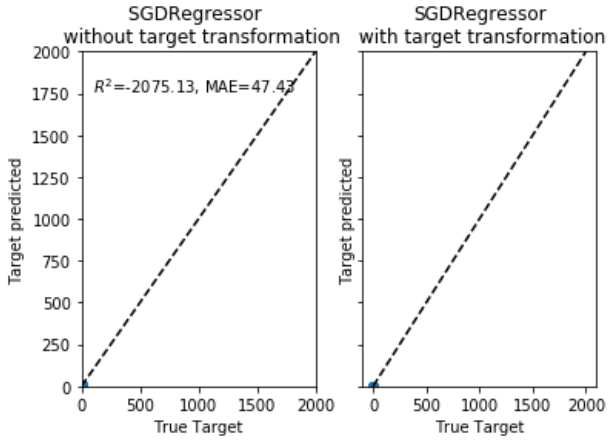

Fig. 12. Feature Scaled SGD Regressor with Target transformation

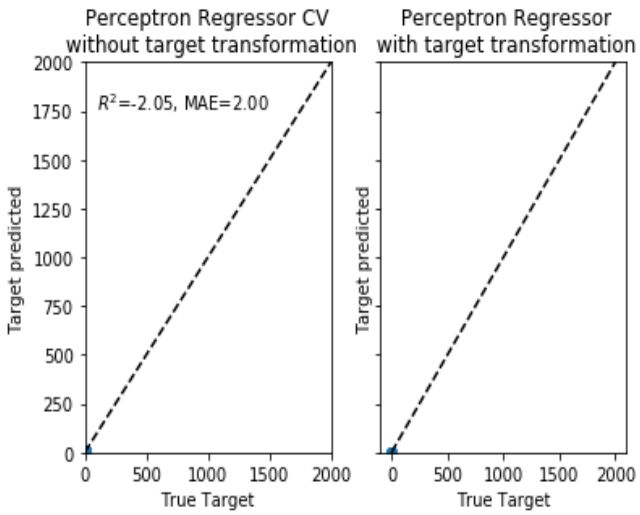

Fig. 13. Perceptron Regressor with Target transformation

The 2 component PCA reduced dataset is fitted to various solvers of logistic regression classifiers like liblinear, lbfgs, newtoncg, sag, saga and warn and their performance is analyzed with the confusion matrix and is shown in Fig. 12.

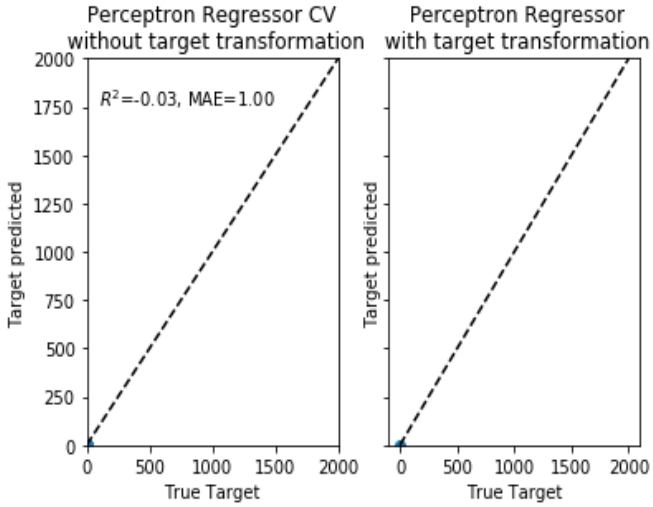

Fig. 14. Feature Scaled Perceptron with Target transformation

The performance analysis for the dataset with and without feature scaling along with the target transformation is shown in Table. 1 – Table 4.

Table. 1 Analysis of Regressors without Transformation

Regressor	No Feature Scaling without Target	
	MAE	R2 Score
Ridge	0.73	0.29
Huber	0.73	0.24
SGD	5.89	0.96
Perceptron	2.05	2.00

Table. 2 Analysis of Regressors with Transformation

Regressor	No Feature Scaling with Target	
	MAE	R2 Score
Ridge	0.71	0.24
Huber	0.73	0.19
SGD	5.86	0.99
Perceptron	2.02	2.01

Table. 3 Analysis of Regressors with Feature Scaling and without Transformation

Regressor	With Feature Scaling without Target	
	MAE	R2 Score
Ridge	0.71	0.24
Huber	44.64	18.43
SGD	2.07	47.43
Perceptron	1.00	0.03

Table. 4 Analysis of Regressors with Feature Scaling and Transformation

Regressor	With Feature Scaling with Target	
	MAE	R2 Score
Ridge	0.71	0.24
Huber	0.73	0.19
SGD	2.04	47.42
Perceptron	1.00	0.04

V. CONCLUSION

This paper attempts to analyze the performance of the regressors with and without applying feature scaling and target transformation. The metrics like mean absolute error and r2 score is analyzed for the dataset before and after applying feature scaling. Experimental results show that, the Perceptron regressor CV has the effectiveness with the mean absolute error of 1.00 and R2 score of 0.04 for the heart disease prediction.

REFERENCES

1. Petkovski, Aleksandar & Risteska Stojkoska, Biljana & Trivodaliev, Kire & Kalajdziski, Slobodan. "Analysis of churn prediction: A case study on telecommunication services in Macedonia". 1-4. 10.1109/TELFOR.2016.7818903, 2016.
2. Jadhav, Rahul & Pawar, Usharani. "Churn Prediction in Telecommunication Using Data Mining Technology", International Journal of Advanced Computer Sciences and Applications. 2. 10.14569/IJACSA.2011.020204, 2011.
3. Hung, Shin-Yuan & Yen, David & Wang, Hsiu-Yu."Applying data mining to telecom chum management. Expert Systems with Applications" . 31. 515-524. 10.1016/j.eswa.2005.09.080, 2006.
4. M. Shyamala Devi, Shakila Basheer, Rincy Merlin Mathew, "Exploration of Multiple Linear Regression with Ensembling Schemes for Roof Fall Assessment using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, October 2019.
5. Shakila Basheer, Rincy Merlin Mathew, M. Shyamala Devi, "Ensembling Coalesce of Logistic Regression Classifier for Heart Disease Prediction using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, October 2019, pp. 127-133.
6. Rincy Merlin Mathew, M. Shyamala Devi, Shakila Basheer," Exploration of Neighbor Kernels and Feature Estimators for Heart Disease Prediction using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, October 2019, pp. 597-605.
7. M. Shyamala Devi, Shefali Dewangan, Satwat Kumar Ambashta, Anjali Jaiswal, Nariboyena Vijaya Sai Ram, "Backward Eliminated Formulation of Fire Area Coverage using Machine Learning Regression", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, October 2019, pp.1565-1569
8. M. Shyamala Devi, Ankita Shil, Prakhar Katyayan, Tanmay Surana, "Constituent Depletion and Divination of Hypothyroid Prevalance using Machine Learning Classification", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.12, October 2019, pp. 1607-1612
9. M. Shyamala Devi, Shefali Dewangan, Satwat Kumar Ambashta, Anjali Jaiswal, Sairam Kondapalli, "Recognition of Forest Fire Spruce Type Tagging using Machine Learning Classification", International Journal of Recent Technology and Engineering, Volume-8 Issue-3, pp. 4309 – 4313, 16 September 2019.
10. M. Shyamala Devi, Usha Vudatha, Sukriti Mukherjee, Bhavya Reddy Donthiri, S B Adhiyan, Nallareddy Jishnu, " Linear Attribute Projection and Performance Assessment for Signifying the Absenteeism at Work using Machine Learning", International Journal of Recent Technology and Engineering, Volume-8 Issue-3, pp. 1262 – 1267, 16 September 2019.
11. M. Shyamala Devi, Mothe Sunil Goud, G. Sai Teja, MallyPally Sai Bharath, "Heart Disease Prediction and Performance Assessment through Attribute Element Diminution using Machine Learning", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no.11, pp. 604 – 609, 30 September 2019.
12. M. Shyamala Devi, Rincy Merlin Mathew, R. Suguna, " Regressor Fitting of Feature Importance for Customer Segment Prediction with Ensembling Schemes using Machine Learning", International Journal of Engineering and Advanced Technology, Volume-8 Issue-6, pp. 952 – 956, 30 August 2019.
13. R. Suguna, M. Shyamala Devi, Rincy Merlin Mathew, "Integrating Ensembling Schemes with Classification for Customer Group Prediction using Machine Learning", International Journal of Engineering and Advanced Technology, Volume-8 Issue-6, pp. 957 – 961, 30 August 2019.
14. Rincy Merlin Mathew, R. Suguna, M. Shyamala Devi, "Composite Model Fabrication of Classification with Transformed Target Regressor for Customer Segmentation using Machine Learning", International Journal of Engineering and Advanced Technology, Volume-8 Issue-6, pp. 962 – 966, 30 August 2019.
15. 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.
16. 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.
17. R.Suguna, M. Shyamala Devi, Rupali Amit Bagate, 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: 10.1080/09720510.2019.1609729ISSN: 0972-0510 (Print), 2169-0014 (Online).
18. R.Suguna, M. Shyamala Devi, Rupali Amit Bagate, 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: 10.1080/09720510.2019.1609729ISSN: 0972-0510 (Print), 2169-0014 (Online).
19. Shyamala Devi Munisamy, Suguna Ramadass Aparna Joshi, "Cultivar Prediction of Target Consumer Class using Feature Selection with Machine Learning Classification", Learning and Analytics in Intelligent Systems, LAIS, Springer, vol. 3, pp. 604-612, June 2019.
20. Suguna Ramadass, 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 entitled "Learning and Analytics in Intelligent Systems, Springer, LAIS vol. 3, pp. 613-620, June 2019.
21. M. Shyamala Devi, Rincy Merlin Mathew, R. Suguna, "Attribute Heaving Extraction and Performance Analysis for the Prophesy 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.
22. R. Suguna, M. Shyamala Devi, 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.
23. Kota Venkateswara Rao, M. Shyamala Devi," Interrogation of Sentiment Perusing with Hash Counting Vectorizer and Term Inverse Frequency Transformer using Machine Learning Classification", International Journal of Recent Technology and Engineering, Volume-8 Issue-4, , 30 November 2019, pp. 3895 - 3901.
24. M. Shyamala Devi, Jyotikinkar Saharia, Shubham Kumar, Aayushi Chansoriya, Prashant Yadav, "Machine Learning Based Suspicion of Customer Detention in Banking with Diverse Solver Neighbors and Kernels", International Journal of Recent Technology and Engineering, Volume-8 Issue-4, , 30 November 2019, pp. 3244 - 3249.
25. M. Shyamala Devi, G. Bhargava Krishna, K.Sowmya, T. Sabari Pavan," Postulation of Customer Retention in Banking Sector using Machine Learning and Principal Component", International Journal of Recent Technology and Engineering, Volume-8 Issue-4, , 30 November 2019, pp. 3178-3182.

26. M. Shyamala Devi, Nariboyena Vijaya Sai Ram, Aravapalli Sai Vamshi, Basyam Bharath, Mallangi Surya Prakash Reddy, "Swindling Shonky Anatomization of Credit Card Transactions using Machine Learning" International Journal of Recent Technology and Engineering, Volume-8 Issue-4, , 30 November 2019, pp. 1477- 1483.
27. M. Shyamala Devi, Kamma Rahul, Manubolu Satheesh, Koruprolu Rajasekhar, Pittala Ganesh Kumar," Count Vectorized Spam and Ham Discernment of Short Message Service using Machine Learning Classification", International Journal of Recent Technology and Engineering, Volume-8 Issue-4, 30 November 2019, pp. 557- 561
28. Kota Venkateswara Rao, M. Shyamala Devi," Disquisition of Sentiment Inquiry with Hashing and Counting Vectorizer using Machine Learning Classification", International Journal of Innovative Technology and Exploring Engineering, vol. 9, no.1, 10 November 2019, pp. 737-743.