International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Backward Eliminated Formulation of Fire Area Coverage using Machine Learning Regression

M. Shyamala Devi, Shefali Dewangan, Satwat Kumar Ambashta, Anjali Jaiswal, Nariboyena Vijaya Sai Ram

Abstract: ***In today's modern world, the environmental wealth is degraded due to the advancement in the technology. The software development leads to the emission of electronic waste that affects the whole part of the country. The forest area and the agricultural land are converted into living places, companies, industries, server warehouse and heavy workstations. Due to this, heavy damages occur in the environmental resources. The basic characteristics of the nature quality are becoming poor due to the technological advancement. Due to the heavy emission of rays in the environment, there is a chance for the occurrence of fire in the forest. This leads to the challenging issue of predicting the area coverage of the fire in the forest. After the event of fire damage, it is a difficult task to analyze the area that suffered from the fire. With this analytical view, this paper focuses on finding the area coverage of fire using various regression algorithms. The forest area coverage dataset from the UCI machine learning repository is used for analyzing the area coverage of the fire. The prediction of area coverage of fire is accomplished in four ways. Firstly, the raw data set is fitted with various regression algorithms to predict the fire area coverage. Secondly, the data set is tailored by the feature selection algorithm namely backward elimination technique. Thirdly, the backward eliminated reduced fire area coverage data set is fitted with various regression algorithms to predict the fire area coverage. Fourth, the performance analysis is done for the raw data set and backward eliminated reduced fire area coverage data set by reviewing the performance metrics mean squared error (MSE), Mean Absolute Error (MAE) and R2 Score. This paper is implemented by python scripts in Anaconda Spyder Navigator. Experimental Result shows that the Passive Aggressive regressor have the effective prediction of fire area coverage with minimum MSE of 0.07, MAE of 1.03 and equitable R2 Score of 0.93 without backward elimination. In the same way, the Passive Aggressive regressor MSE of 0.06, MAE of 1.02 and equitable R2 Score of 0.96 with backward elimination.***

Index Terms: Machine Learning, Feature Extraction, PCA, MSE, MAE, R2 Score.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

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.

Shefali Dewangan, II Year B.Tech Student, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

Satwat Kumar Ambastha, II Year B.Tech Student, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

Anjali Jaiswal, II Year B.Tech Student, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

Nariboyena Vijaya Sai Ram, III Year B.Tech Student, Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, TamilNadu, India.

I. INTRODUCTION

The machine learning technology is used for the prediction and the forecasting of the damages that can occur with the natural resources by the knowledge of the occurred event. Generally, the damages occurred by the natural calamity must be interpreted and analyzed numerically for the future prediction of the damages. The world exists only because of the availability of natural resources and its wealth. The living organisms are available only by the presence of agricultural land, forest and other natural resources. So, the technological advancement should favor the world by predicting the damages that can occur due to the natural calamity. In machine learning technology, the prediction of dependent attribute is performed using regression or classification.

The paper is equipped in which the review of literatures is discussed with Section 2 go subsequently by the proposed work in the Section 3. Implementation and Performance Analysis is deliberated in Section 4 tailed by the conclusion of the paper in Section 5.

II. RELATED WORK

A. Literature Review

SAMOA is used to analyze the forest fire storm using the machine learning algorithms in SAMOA which is similar to Mahout for Hadoop. Vertical Hoeffding parallelizing streaming decision tree induction is encompassed in SAMOA API procedure for forecasting the forest cover types from the cartographic variables [1].

The mapping of the abstract dataset variables and the interconnected components are the basic requirements for the prediction of forest design and its health. The one to one mapping of the forest attributes and their components are done for northern Europe by using k-nearest neighbor's method for augmenting the wall-to-wall basal area, volume, and cover type assessment. Some of the factors like distance metric, weighting function, feature weighting parameters, and number of neighbors are also measured for extrapolation of the forest cover type [2].

Resources Renewable Planning Act (RPA) is also used for finding the area coverage of the fire and for the mapping of forest attributes and its components [3].

A multi-source account process for the forest area coverage is established grounded on the k-nearest neighbor rule for generating the maps of nominated forest variables.

It uses pulverized proposal from the NFI, satellite image data and pounded cloak maps from Landsat data[4]. The label noise is one of the major trepidations in classification. The review on label noise and its influence on classification is done by using label noise-robust, label noise cleansing, and label noise-tolerant algorithms [5].

The machine learning attribute collection and removal methods can be used for the detection of dependent attribute for various real time application can be understood through this article [6]–[17].

III. DIMENSIONALITY REDUCTION

Dimensionality reduction accomplishes the conversion of extraordinary dimensional data to lesser dimensional data deprived of any damage of information. There are two group in dimensionality reduction: Feature selection and Feature Extraction. Feature selection look for at preservation of only suitable features in the given data set whereas feature extraction concentrates on finding an optimal set of attributes usually a combination of input attribute without loss in the originality of information.

A. Backward Elimination

The stages in the Backward Elimination are as follows,
Step 1: Predefine the significance level to quit the model at the end of the processing.
Step 2: Tey to build and design the model with the best independent attributes.
Step 3: Examine the independent attribute with highest P value.
Step 4: If P value is less than the significance level then go to step 6.
Step 5: Is step 4 is not satisfied, then the "model is stabilized to use".
Step 6: Eliminate the independent variable

IV. PROPOSED WORK

In this work, the area coverage of the fire is predicted by using the machine learning algorithms. Our contribution in this paper is folded in two ways.

(i) Firstly, the raw data set is fitted with various regression algorithms to predict the fire area coverage and they are as follows,
- Linear Regression
- ARDRegression Module
- Bayesian Ridge
- Huber Regressor
- LarsCV
- LassoCV
- LassoLarsCV
- OrthogonalMatchingPursuit
- PassiveAggressiveRegressor
- RANSAC Regressor
- RidgeCV
- SGDRegressor
- TheilSenRegressor
- ElasticNet
- ElasticNetCV
- LassoLarsIC

(ii) Secondly, the data set is tailored by the feature selection algorithm namely backward elimination technique.

(iii) Thirdly, the backward eliminated reduced fire area coverage data set is fitted with above mentioned regression algorithms to predict the fire area coverage.

(iv) Fourth, the performance analysis is done for the raw data set and backward eliminated reduced fire area coverage data set by reviewing the performance metrics mean squared error (MSE), Mean Absolute Error (MAE) and R2 Score.

A. System Architecture

The overall design of our work is shown in Fig. 1

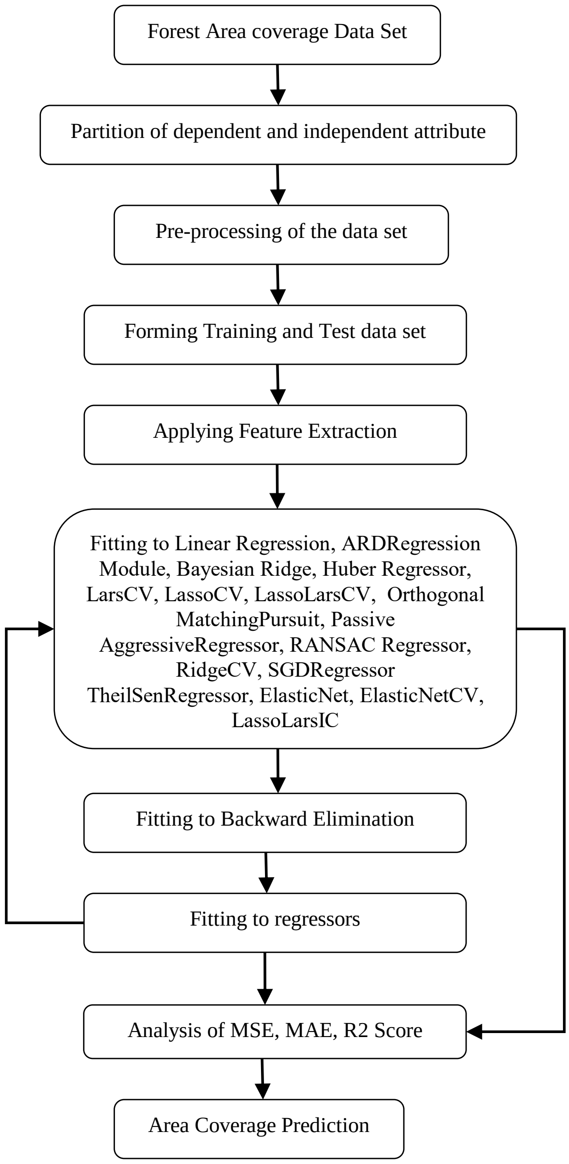

Fig. 1 System Architecture

V. IMPLEMENTATION AND PERFORMANCE ANALYSIS

A. Forest Area Coverage Prediction

The Forest Area Coverage dataset from UCL machine learning Repository is used for implementation with 1910 independent attribute and 1 Forest Area Coverage dependent attribute. The attribute are shown below.

1) X
2) Y
3) FFMC - Fine Fuel Moisture Code
4) DMC – Duff Moisture Code
5) DC - Drought Code
6) ISI - Initial Spread Index
7) Temp
8) RH - Relative humidity
9) Wind
10) Rain
11) Area - Dependent Attribute

The relationship and the probability of matching with each of the attributes in the forest area coverage is shown in Fig 2.

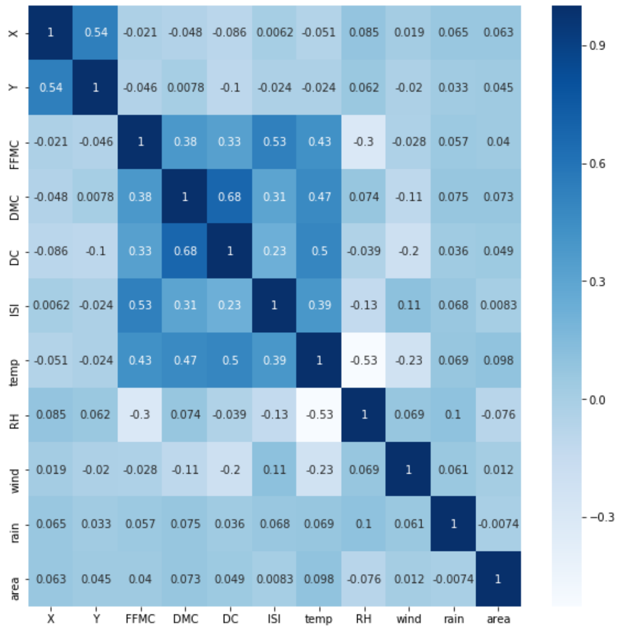

Fig. 2 Relationship of Dataset

The dataset is subjected to backward elimination and the first step is shown in Fig 3 – Fig. 9.

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.2384	47.505	0.068	0.946	-90.151	96.627
x1	1.3547	1.135	1.193	0.233	-0.877	3.587
x2	1.0347	2.158	0.480	0.632	-3.207	5.277
x3	-0.0981	0.491	-0.200	0.842	-1.064	0.868
x4	0.1021	0.053	1.921	0.055	-0.002	0.207
x5	-0.0125	0.013	-0.963	0.336	-0.038	0.013
x6	-0.3067	0.579	-0.530	0.597	-1.445	0.832
x7	0.4124	0.635	0.649	0.517	-0.837	1.661
x8	-0.1783	0.189	-0.943	0.346	-0.550	0.194
x9	0.9097	1.319	0.690	0.491	-1.683	3.502
x10	-2.0101	6.895	-0.292	0.771	-15.566	11.545

Fig. 3 First Step in backward elimination

	coef	std err	t	P>|t|	[0.025	0.975]
const	4.4990	23.145	0.194	0.846	-41.001	49.999
x1	4.4990	23.145	0.194	0.846	-41.001	49.999
x2	1.6320	0.955	1.709	0.088	-0.245	3.509
x3	-0.1141	0.489	-0.233	0.816	-1.076	0.848
x4	0.1052	0.053	1.995	0.047	0.002	0.209
x5	-0.0130	0.013	-1.004	0.316	-0.038	0.012
x6	-0.3076	0.578	-0.532	0.595	-1.443	0.828
x7	0.3802	0.626	0.607	0.544	-0.851	1.611
x8	-0.1895	0.186	-1.020	0.308	-0.555	0.176
x9	0.8630	1.310	0.659	0.510	-1.712	3.438

Fig. 4 Second Step in backward elimination

	coef	std err	t	P>|t|	[0.025	0.975]
const	3.6886	23.164	0.159	0.874	-41.847	49.224
x1	3.6886	23.164	0.159	0.874	-41.847	49.224
x2	2.3691	1.817	1.304	0.193	-1.203	5.941
x3	-0.0767	0.490	-0.157	0.876	-1.039	0.886
x4	0.1017	0.053	1.931	0.054	-0.002	0.205
x5	-0.0142	0.013	-1.102	0.271	-0.039	0.011
x6	-0.1976	0.561	-0.352	0.725	-1.301	0.905
x7	0.3134	0.607	0.516	0.606	-0.879	1.506
x8	-0.1819	0.184	-0.987	0.324	-0.544	0.181

Fig.5 Third Step in backward elimination

	coef	std err	t	P>|t|	[0.025	0.975]
const	-5.5333	19.567	-0.283	0.777	-43.998	32.932
x1	-5.5333	19.567	-0.283	0.777	-43.998	32.932
x2	1.5195	0.950	1.599	0.110	-0.348	3.387
x3	0.0292	0.471	0.062	0.951	-0.897	0.956
x4	0.0863	0.048	1.784	0.075	-0.009	0.181
x5	-0.0151	0.013	-1.180	0.239	-0.040	0.010
x6	-0.2843	0.557	-0.510	0.610	-1.379	0.811
x7	0.6831	0.480	1.424	0.155	-0.260	1.626

Fig. 6 Fourth Step in backward elimination

	coef	std err	t	P>|t|	[0.025	0.975]
const	-2.8520	12.952	-0.220	0.826	-28.313	22.609
x1	-2.8520	12.952	-0.220	0.826	-28.313	22.609
x2	-2.8520	12.952	-0.220	0.826	-28.313	22.609
x3	0.1731	0.463	0.374	0.708	-0.736	1.082
x4	0.0941	0.048	1.948	0.052	-0.001	0.189
x5	-0.0117	0.012	-0.955	0.340	-0.036	0.012
x6	-0.1078	0.547	-0.197	0.844	-1.184	0.968

Fig. 7 Fifth Step in backward elimination

	coef	std err	t	P>|t|	[0.025	0.975]
const	-6.0980	12.448	-0.490	0.624	-30.569	18.373
x1	-6.0980	12.448	-0.490	0.624	-30.569	18.373
x2	-6.0980	12.448	-0.490	0.624	-30.569	18.373
x3	2.4893	1.800	1.383	0.167	-1.050	6.028
x4	0.1215	0.413	0.294	0.769	-0.690	0.933
x5	0.0623	0.036	1.716	0.087	-0.009	0.134

Fig.8 Sixth Step in backward elimination

	coef	std err	t	P>|t|	[0.025	0.975]
const	-9.9458	11.740	-0.847	0.397	-33.023	13.132
x1	-9.9458	11.740	-0.847	0.397	-33.023	13.132
x2	-9.9458	11.740	-0.847	0.397	-33.023	13.132
x3	1.4846	0.946	1.569	0.117	-0.375	3.344
x4	0.3679	0.384	0.958	0.339	-0.387	1.123

Fig. 9 Seventh Step in backward elimination

B. Analysis with Backward Elimination

The raw data set is fitted with various regression algorithms Fitting to Linear Regression, ARDRegression Module, Bayesian Ridge, Huber Regressor, LarsCV, LassoCV, Lasso LarsCV, Orthogonal Matching Pursuit, Passive Aggressive Regressor, RANSAC Regressor, RidgeCV, SGD Regressor, TheilSen Regressor, Elastic Net, ElasticNetCV regressor, LassoLarsIC to predict the fire area coverage. The performance analysis is done with MAE, MSE and R2 Score. The metric comparison of MAE for the forest fire area coverage dataset is shown in Table 1 and the analysis is shown in Fig 10.

Table. 1. Comparison of Mean Absolute Error

Regression Methods	Mean Absolute Error	
	Without Backward Elimination	With Backward Elimination
Linear	1.45	1.41
ARDModule	1.64	1.61
Bayesian Ridge	1.75	1.73
Huber	1.61	1.57
LarsCV	1.81	1.73
LassoCV	1.79	1.70
LassoLarsCV	1.73	1.64
Orthogonal	1.63	1.51
Passive Aggressive	1.03	1.02
RANSAC	1.43	1.31
RidgeCV	1.59	1.48
SGDRegressor	1.61	1.56
TheilSen	1.68	1.54
ElasticNet	1.72	1.61
ElasticNetCV	1.32	1.25
LassoLarsIC	1.29	1.22

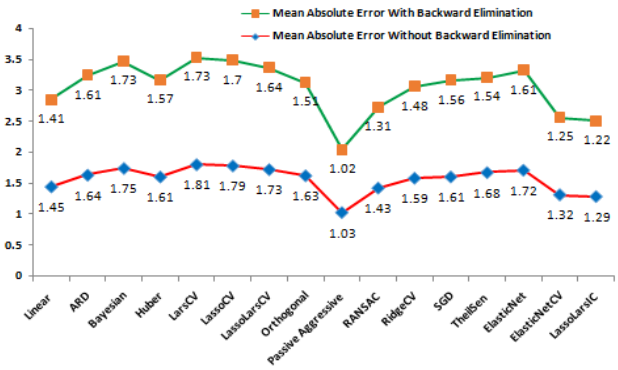

Fig. 10 Analysis of Mean Absolute Error

The metric comparison of MSE for the forest fire area coverage dataset is shown in Table 2 and the analysis is shown in Fig 11.

Table. 2. Comparison of Mean Squared Error

Regression Methods	Mean Squared Error	
	Without Backward Elimination	With Backward Elimination
Linear Regression	0.29	0.24
ARDRegression	0.28	0.22
Bayesian Ridge	0.32	0.39
Huber Regressor	0.28	0.25
LarsCV	0.27	0.24
LassoCV	0.22	0.17
LassoLarsCV	0.17	0.17
Orthogonal	0.18	0.15
Passive	0.07	0.06
RANSAC	0.21	0.18
RidgeCV	0.37	0.31
SGDRegressor	0.34	0.27
TheilSen Regressor	0.29	0.23
ElasticNet	0.26	0.21
ElasticNetCV	0.19	0.15
LassoLarsIC	0.33	0.23

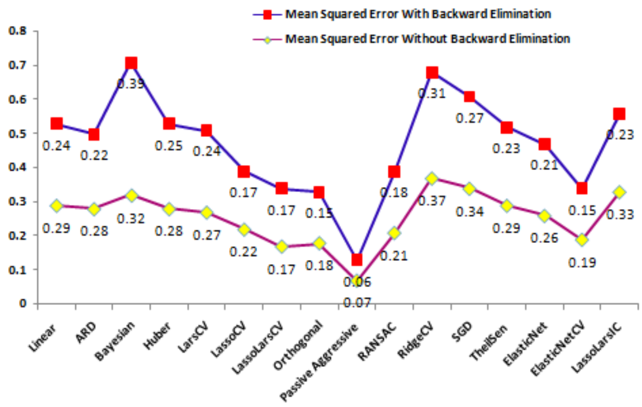

Fig. 11 Analysis of Mean Squared Error

The metric comparison of R2 Score for the forest fire area coverage dataset is shown in Table 3 and the analysis is shown in Fig 12.

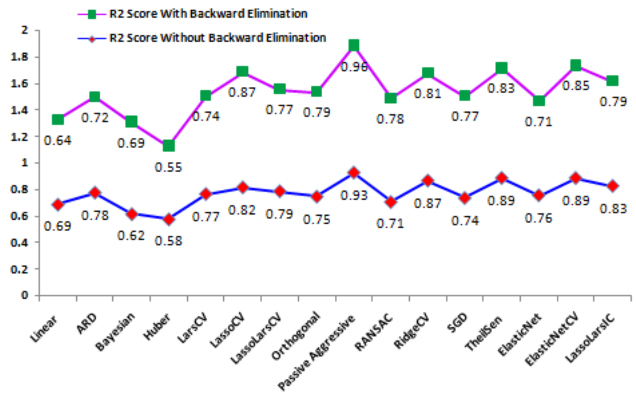

Fig. 12 Analysis of R2 Score

Table. 2. Comparison of R2 Score

Regression Methods	R2 Score	
	Without Backward Elimination	With Backward Elimination
Linear	0.69	0.64
ARDRegression	0.78	0.72
Bayesian Ridge	0.62	0.69
Huber	0.58	0.55
LarsCV	0.77	0.74
LassoCV	0.82	0.87
LassoLarsCV	0.79	0.77
Orthogonal	0.75	0.79
Passive	0.93	0.96
RANSAC	0.71	0.78

RidgeCV	0.87	0.81
SGDRegressor	0.74	0.77
TheilSen	0.89	0.83
ElasticNet	0.76	0.71
ElasticNetCV	0.89	0.85
LassoLarsIC	0.83	0.79

The iterations that are carried out in the backward elimination that are applied in the forest fire area coverage dataset is shown with the attribute details and the attributes that are eliminated and removed in each iteration is shown in the Table 4.

Table. 4. Iterations of Backward Elimination

Features	Iterations of Backward Elimination						
	I	II	III	IV	V	VI	VII
X	0.94	---	---	---	---	---	---
Y	0.23	0.84	---	---	---	---	---
FFMC	0.63	0.84	0.87	0.77	0.82	0.62	0.03
DMC	0.84	0.88	0.87	0.77	0.82	0.62	0.03
DC	0.05	0.81	0.19	0.11	0.82	0.62	0.03
ISI	0.33	0.04	0.87	---	---	---	---
Temp	0.59	0.31	0.05	0.95	---	---	---
RH	0.51	0.59	0.27	0.07	0.70	0.16	0.01
Wind	0.34	0.54	0.72	0.23	0.05	0.76	X
Rain	0.49	0.30	0.60	0.61	0.34	0.08	0.03
Area	0.77	0.51	0.32	0.15	0.84	---	---

VI. CONCLUSION

This paper analyses the performance of forest area coverage with and without backward elimination. The performance analysis is done with the metrics such as MAE, MSE and R2 Score. Experimental Result shows that the Passive Aggressive regressor have the effective prediction of fire area coverage with minimum MSE of 0.07, MAE of 1.03 and equitable R2 Score of 0.93 without backward elimination. In the same way, the Passive Aggressive regressor MSE of 0.06, MAE of 1.02 and equitable R2 Score of 0.96 with backward elimination.

REFERENCES

1. G. D. Li et al., "Forest Cover Types Classification Based on Online Machine Learning on Distributed Cloud Computing Platforms of Storm and SAMOA", Advanced Materials Research., Vol. 955-959, 2014, pp. 3803-3812.
2. H. Franco-Lopez, A. R. Ek and M. E. Bauer, "Estimation and mapping of forest stand density volume and cover type using the k-nearest neighbors method", Remote Sensing of Environment, vol. 77, no. 3, 2001, pp. 251-274.
3. Zhu Z and D. L. .Evans, “US forest types and predicted percent forest cover from AVHRR data” Photogrammetric Engineering and Remote Sensing., Vol. 60(5), 1994, pp. 525-532.
4. B. T. Wilson, A. J. Lister and R. I. Riemann, "A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data", Forest Ecology and Management, vol. 271, 2012, pp. 182-198.
5. 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.
6. 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.
7. 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.
8. 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.
9. M. Shyamala Devi, Rincy Merlin Mathew, and 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.
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. 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.
12. 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, 30 September 2019.
13. 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, 30 September 2019.
14. 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, 30 September 2019
15. 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
16. 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.
17. 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.