Enhancement of Students' Learning Experience on Educational Platform

[1]S. Manohar, [2]Rumi Shaharyar, [3]Kanu Kartikeya, [4]Kavya Manjari

[1]Assistant Professor, [2]Student, [3]Student, [4]Student

[1]Department of Computer Science and Engineering,

[1]SRM Institute of Science and Technology, Chennai, Tamil Nadu, India

Abstract: This paper focuses on improving approaches to better understand students and their functional choices of interest on both digital and institutional (e.g.: schools, colleges, universities etc.) form of education. The exact sequence of CRISP-DM methodology was followed on the data collected from both the platforms separately, including a number of machine learning algorithms to get a more detailed inference view of our model. Since our motive is to improve the predictions and find accurate insights to justify our inferences, LASSO regression was applied and further RIDGE regression to take the verification of our accuracy up a notch. After all this, we can generalize our conclusion by stating that each and every combination of attributes can have a significant effect on the final prediction, resulting in varied areas of interest of students and therefore further classification techniques can be performed in order to produce more singularized results.

Keywords - **crisp-dm; lasso; ridge regression; classification, online learning.**

I. INTRODUCTION

A small sample description of data can be seen in the figure (1.1) shown as a data frame, displaying general statistics of attribute overall percentage and age of students. The data has been collected from both digital platforms as well as through thorough research from high school students and teachers. The data contains attributes in both nominal and ordinal form of different data types integer, float and string. Where in order to maintain accuracy and generate less noise, the categorical variables have been converted to dummy variables and further multiple linear regression in order to reduce our extensive data to a more predictable form.

Algorithms for regression and classification techniques have been applied in a similar fashion as in house price prediction, but with multiple combinations of dependent and independent variables to simplify complexity and generate different inferences according to the statistics observed. The most common conclusion from each observation tells the variability of interests in students does not linearly or proportionally depend on singular relationships, e.g.: comparing statistics of a student completing homework (suppose science) but scoring less in the exam (science) doesn't prove his/her lack of interest.

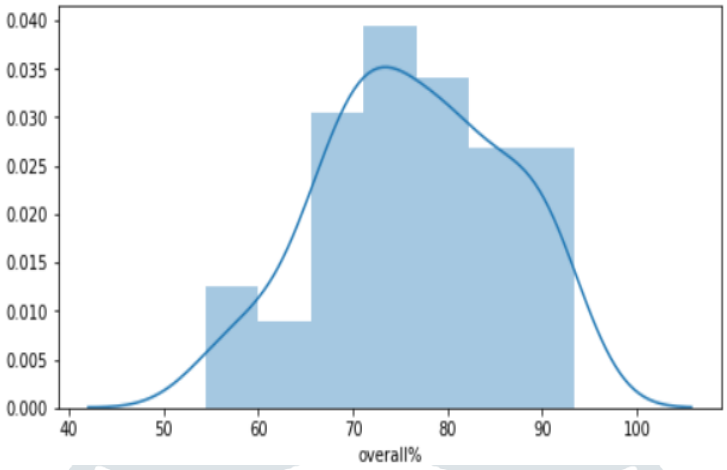
	OVERALL %	AGE
count	100.000000	100.000000
mean	76.408000	13.750000
std	9.765034	1.305582
min	54.400000	12.000000
25%	69.400000	13.000000
50%	75.800000	14.000000
75%	83.300000	15.000000
max	93.400000	17.000000

Figure 1.1

II. EXPLORING THE DATA

The graph below shows an overall % of all the students calculated by a normal distribution in a variate X with mean μ and variance σ^2 is a statistic distribution with probability density function on the domain (x_minimum, x_maximum). From the graph we can observe the most of our data is centralized around the range of 70-80.

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \quad (1)$$



OLS Regression Results

Dep. Variable:	y	R-squared:	0.945
Model:	OLS	Adj. R-squared:	0.943
Method:	Least Squares	F-statistic:	551.5
Date:	Thu, 21 Mar 2019	Prob (F-statistic):	2.30e-60
Time:	19:13:11	Log-Likelihood:	-224.11
No. Observations:	100	AIC:	456.2
Df Residuals:	96	BIC:	466.6
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	21.5491	2.831	7.613	0.000	15.930	27.168
x1	0.4270	0.028	15.367	0.000	0.372	0.482
x2	0.3877	0.028	13.734	0.000	0.332	0.444
x3	-0.1603	0.027	-6.004	0.000	-0.213	-0.107

Omnibus:	11.652	Durbin-Watson:	2.098
Prob(Omnibus):	0.003	Jarque-Bera (JB):	28.080
Skew:	0.270	Prob(JB):	7.99e-07
Kurtosis:	5.539	Cond. No.	1.39e+03

Figure 1.2

The above statistics are output from multiple linear regression, the equation used is y=b0+b1x1+b2x2+...+bnxn where y is the independent variable or target variable or value to be predicted, b0 to bn are the coefficients, x1 to xn are the different attributes. In figure 1.2, const refers to b0 and x1,x2,x3 are the variables incurred after eliminating the least significant attributes(or variables i.e. other values of x) R-squared and Adj. R-squared values define the accuracy of our algorithm, obtained from the formula (2) and then squaring the value of r which also tells us the linearity, strength and direction of the regression line. Here 0.945 means 94.5% accuracy has been obtained from our regression model.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (2)$$

III. DATA PRE-PROCESSING

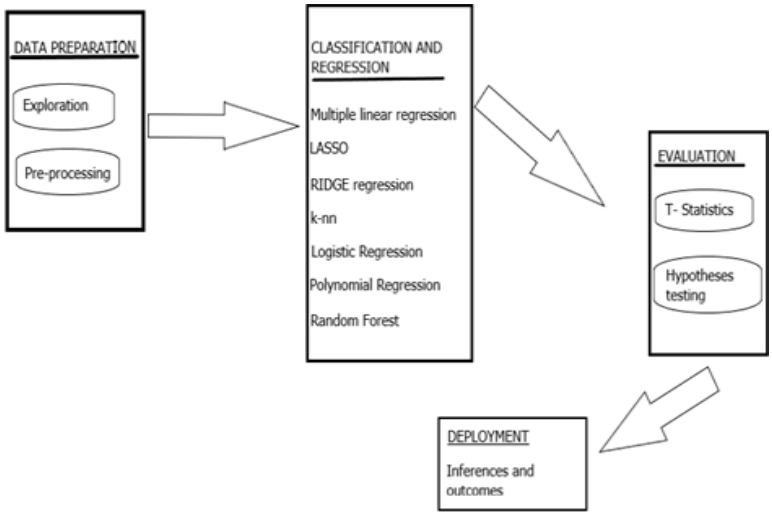


Figure 1.3

Our process of data pre-processing starts with consolidating data, according to our model requirements. Our model requires data for improving student choices and their performance, followed by data cleaning, which involves removing inconsistency and noise from our data, in order to achieve a smoother workflow for our model and precise prediction of target values.

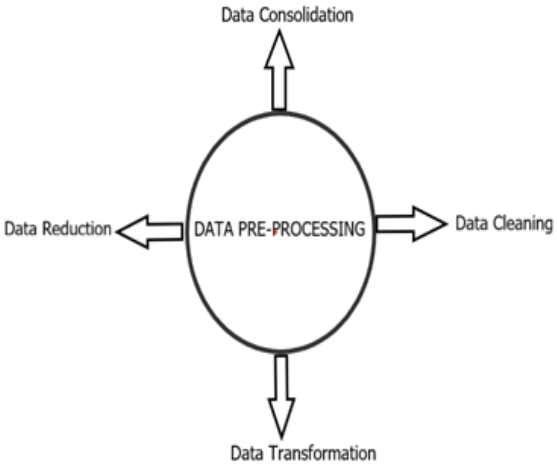


Figure 1.4

Data transformation, most necessary step to take before applying the model, in our case it includes maintaining the consistency of categorical variables by creating dummy variables for the respective categories i.e. which are of 0 and 1 form where 0 means false and 1 means true. Data transformation is directly linked to data reduction which involves us to select our features, with a technique referred as feature selection. Feature selection and scaling, generally scaling are performed on normally distributed data and need to be standardized if not normally distributed. This feature scaling is done in different forms like, rescaling data where the whole dataset is rescaled to a range between 0 and 1, standardizing data where by-default each attribute has a mean of 0 and standard deviation of, binarizing data where we transform data using a binary threshold in which all values above the threshold are marked 1 and all equal to or below are marked as 0.

IV. CLASSIFICATION AND REGRESSION

Observing the pattern and the output from multiple linear regression, we could understand that a simple linear fit won't be an accurate method to predict our data with higher r-squared values, therefore we perform polynomial random forest in order to get better fit for the data and improve our accuracy.

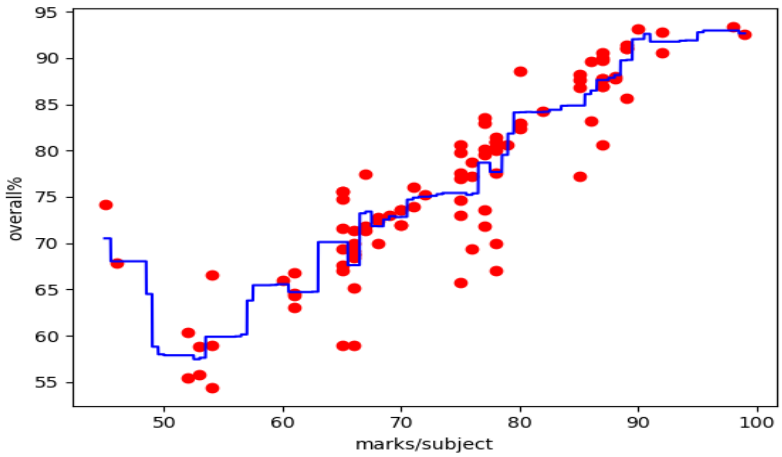


Figure 1.6

The graph figure (1.6), shows a line that fits well with the data and is based on the random forest which basically is performing both regression and classification tasks with the use of multiple decision trees, here estimating our coefficient that determines the total decision trees is equal to 1000. Here, we have taken an instance of a pair of combination obtained after performing ridge regression guiding us to form pairs best needed to make predictions of our target value overall %. Before achieving random forest regression, it was made sure of its importance that is how useful could it be to us by performing SVR support vector regression which is responsible to tell us whether an approach towards polynomial regression will increase our accuracy or rather decreases. So, from SVR our resultant form was polynomial regression.

V. EVALUATION

Evaluation depends on factors like maintaining high accuracy, without letting our model overfit or underfit, rejection and not a rejection of our hypotheses and further calculating our T-statistics because only all these factors' evaluation can help us to determine the success of the model.

Hypotheses testing refers to generating our null hypotheses (Ho) and alternate hypotheses ($H\propto$) and based on the p-value and our significance level (set here as 0.04) we will conclude whether to reject or not reject our hypotheses. That is p-value >0.04 will not reject our Ho whereas p-value < 0.04 will reject our hypotheses hence proving our assumptions and expectations wrong.

REFERENCES

[1] Sangho Suh, Computer Science, Korea University
Seoul, Republic of Korea, "Learning Analytics & Educational Data Mining."

[2] Tut Herawan, Ashish Dutt, and Maizatul Akmar Ismail, "A systematic review on educational data mining", 2016.
[3] K. Z. Aung and N. N. Myo, "Sentiment analysis of students' comment using lexicon-based approach," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, 2017.
[4] G. Siemens and R. S. J. D. Baker, ''Learning analytics and educational data mining: Towards communication and collaboration,'' presented at the 2nd Int. Conf. Learn. Anal. Knowl. (LAK).
[5] M. M. A. Tair and A. M. El-Halees, ''Mining educational data to improve students' performance: A case study,'' Int. J. Inf. Commun. Technol. Res., vol. 2, no. 2, pp. 1–7, Feb. 2012.
[6] C. Romero, S. Ventura, and E. García, ''Data mining in course management systems: Moodle case study and tutorial,'' Comput. Edu., vol. 51, no. 1, pp. 368–384, Aug. 2008.
[7] B. Azarnoush, J. M. Bekki, G. C. Runger, B. L. Bernstein, and R. K. Atkinson, ''Toward a framework for learner segmentation,'' J. Educ. Data Mining, vol. 5, no. 2, pp. 102–126, 2013.
[8] S. K. Mohamad and Z. Tasir, ''Educational data mining: A review,'' Proc.-Social Behavioral Sci., vol. 97, pp. 320–324, Nov. 2013.
[9] J. C. Turner, P. K. Thorpe, and D. K. Meyer, ''Students' reports of motivation and negative affect: A theoretical and empirical analysis,''J. Educ. Psychol., vol. 90.
[10] C. Romero, M. I. López, J.-M. Luna, and S. Ventura, ''Predicting students' final performance from participation in on-line discussion forums,'' Comput. Edu., vol. 68.
[11] A. F. Wise, J. Speer, F. Marbouti, and Y.-T. Hsiao, ''Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviours," Instructional Sci., vol. 41.