TECHNOLOGIES USED IN SCHOOL

Analytical CRM

Name: Arth Vyas
Student ID:X17170516
Subject: Analytical CRM
Professor: Vikas Sahni
Course: MSc. Data Analytics (cohort A)
Submission Date: 03/04/2019

Technologies used in School

Arth Vyas
x17170516
MSc in Data Analytics (Cohort A)
National College of Ireland

Abstract - In this project of Analytical CRM an analysis on technology and tools used in school will be carried out with respect to the student enrolment. This will analyze the approach different variable impacting on the intake process and can analyze the financial improvement. To understand the correlation different technique will be used. The analysis will be boosting the understanding of this traditional process of enrolment and give new dimension to school administration which can be useful to understand the process of enrolment. This report will provide technical approach to view enrolment in different manner.

Keywords— Multiple linear regression, analysis, regression, spss, data mining, efficiency, customer relationship management, multiple linear regression

I. INTRODUCTION

Teaching being an important part of our day to day life, one is continuously in the search of effective way for imparting knowledge. There are different ways for delivering knowledge to a student since from childhood till achieving highest degree. Everybody wants to pursue their degree from honored organization or willing to enroll their child in reputed school or university for the betterment and smart growth of one. In this journey for preparing the students for the coming future the approach considered by school or university is important as well as how they are utilizing the latest technologies to transfer knowledge in much more interactive way to students. So, the tools and technology provided to a student plays a vital roll in the development of them and on the other hand it also gives ease to parents or student to search the better school to study or get enroll in it for gaining knowledge. Keeping in mind it can be seen that schools, university and colleges are focusing on tools and technology to provide best education for the improvement and smart learning. Nowadays this requirement is key area which are often considered as priority before selecting particular schools/college for enrollment of their children.

A. Dataset

The Dataset selected for the analysis is about the tools and technology or we can say facilities provided in a school for student. A survey has been carried out for the several field mentioned above which can be analyzed accordingly or as required. Some of the attributes present in the dataset are.

- School
- District
- Year
- Enrollment
- Advance classroom
- Desktop
- Laptop
- Tablet
- Student email account
- Cameras
- Number of Projector
- Class with internet access
- Screen resolution

The data contains of approximately 2800 rows and numerical data. The data has been taken from Data.Gov.com, the published date 02/01/2018. Below is the link of dataset
https://catalog.data.gov/dataset/school-district-technology-survey

Data Preprocessing

To carry out Multiple linear regression the downloaded data was taken for the first step which is Data preprocessing here the unwanted rows were deleted, and column named were defined as required. Once the data is ready for the analysis is then imported into SPSS tool to perform linear multiple regression model. Here in this dataset **Enrollment** is taken as dependent variable

II. INVESTIGATION OF SUITABLE TECHNIQUES

The dataset has been analyzed in the Rapid miner to understand normal correlation between each field and to know which type of data it is and what is appropriate model to carry out analysis.

The rapid miner analysis was carried out as follows. First data has been imported into the rapid miner, then a target has been set and after than a dependent variable has been set against which all the variable will be checked that if they have the correlation with each other. According to the rapid miner analysis and the looking at the nature of the data it seems to be analyze it in a regression model. It is the model where a correlation is present or not in between variable out which one need to select a dependent variable as well as independent variable which in turn will give statistical proof that relation exist between the variable.

.

III. LITERATURE REVIEW

A approach has been taken to identify the enrollment of student decisions is better done in online courses of college success rate as well as the student who get enroll by face-to-face [1]. This study used a sample data of 24,006 student to validate and check the reliability on online readiness survey. A multilevel model were applied to check the predictor is true or not in online vs face-to-face course outcome irrespective of complexity of course taken by student. It was found that online readiness survey may do attract the student to get enroll even though there is no risk of applying online. Student tends to get enroll face-to-face more than the online. If college find that the existing readiness survey of student at risk in academic environment than alternative method

should be brought by college. Characteristic obtained from research department that identify the group of student and provide extra support online, advising, tutoring. Hence, screening method should be rigorously thought and implemented accordingly.

A analysis of one laptop per child program introduced by the Catalan government is carried to understand whether it help to improve the performance of student [2]. It has been observed that this program had negative impact on students who are learning English and mathematics the test score drastically reduce by0.20-0.22 standardized points which represent 3.8%-6.2% of the average score. It was found more strong effect in boys than girls. Hence in this paper a technology offer by school is checked if it has significant correlation with enrollment or not.

A three step logit model is applied and tested for the child enrolment after completing the primary education and then moving secondary education of Ghana school. Few findings were there are listed as below [3]. It was observed that parental education, household income, gender of head of household plays a vital role in enrollment of student. A decent record was obtained for the student who complete the primary successfully and will be taken forward in secondary. Finally family resources and guidance play important role in enrollment of student.

In this paper overall technology related factors and family resources were consider to find the correlation between the enrollment of student in school. The detailed analysis and methodology is proposed further in this paper.

IV. METHODOLOGY

The methodology is the important part of any project below is the hypotheses developed for this test and as mentioned above the data the appropriate algorithm for analysis will be disclose in the implementation section.

A. *Hypothesis*

H_0: **Tools and technology affect the enrollment of student**

H₁: **Tools and technology does not affect enrollment of student**

V. IMPLEMENTATION

The Multiple linear regression is the one of the methods in regression. The formula of multiple linear regression is shown below [4].

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_{ip} + \varepsilon \text{ for } i = 1,2 \dots n.$$

Here the dependent variable is dependent on multiple independent variable.

To perform this regression a data was first inserted in SPSS the screen shot of data is given below.

Here, before importing the CSV file into SPSS basic extraction, transformation and cleaning was performed on data as explained in data preprocessing section. The multiple linear regression have some assumption which needs to be followed they are linearity, normally distributed, homoscedasticity and multi collinearity.

The data here for linear multiple regression has been inserted as step wise method the screenshot of stepwise selection is given below.
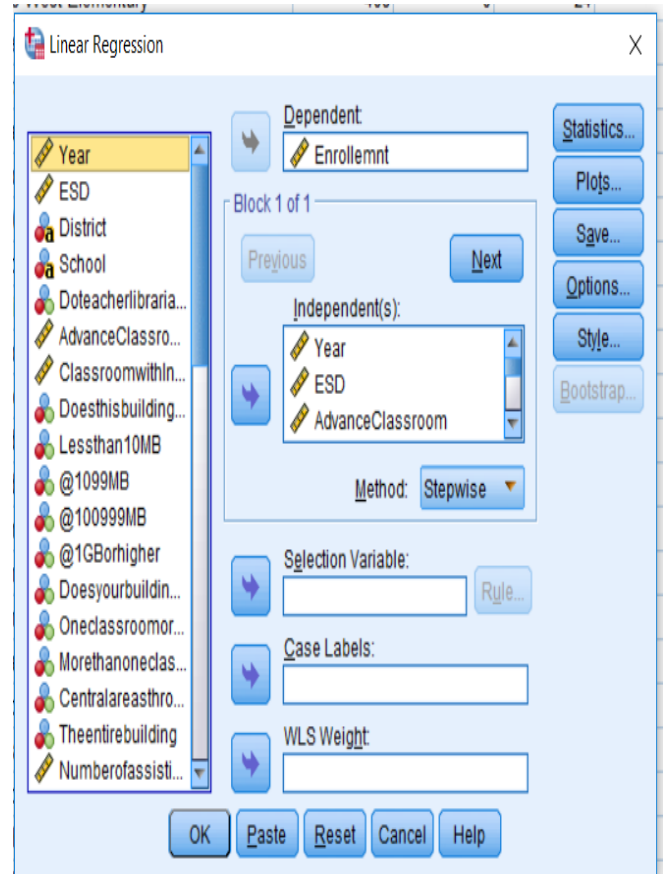
which will facilitate us with the unwanted variable who do not have any correlation or does not fall under statistic criteria. If P value is >= .100 then model will remove the variable which indirectly represent that variable does not have statistical correlation with the dependent variable.

In this dataset there were around 54 variables, once the data entry method has been selected than out of 54 only 20 variables were selected by regression model which make statistically significant correlation between dependent and independent variable.

Interpretation of model and results

First of all let us investigate Coefficient table here three main column will be important for

analysis tolerance value will be check if there is any multicollinearity between variable it should not be less than .10 in our case all the value are above .10. The same is with the VIF value if it is greater than 10 means there is multicollinearity between variables in our case we have 4 variables with multicollinearity which are Number of tablets, Advance classroom, Numbers of IOS, and Number of classrooms with internet access.

Let us go through the Model Summary shown below. Here R sq will tell us how much variance is present for dependent variable which is 92.2 % on the other hand Adjusted R sq value state the same as R sq value but if it has overestimated then

Adjusted R sq will fine tune it, here Adjusted R sq is 88.6 % in short, we can say that we can observe 88.6 % variance in Enrollment through this model.

dependent variable top 5 variable which make strongest unique contribution enrollment are Advance classroom, cameras, screen resolution, email account, and desktop they all have significance value less than 0.5 which means they will make statistically significant contribution to dependent variable. In last partial

correlation of these variable will indicate how much percent affect will be happen on variance in independent variable. 0.25%, 11%, 5%, 6%, and 3.5% for above mentioned independent variable.
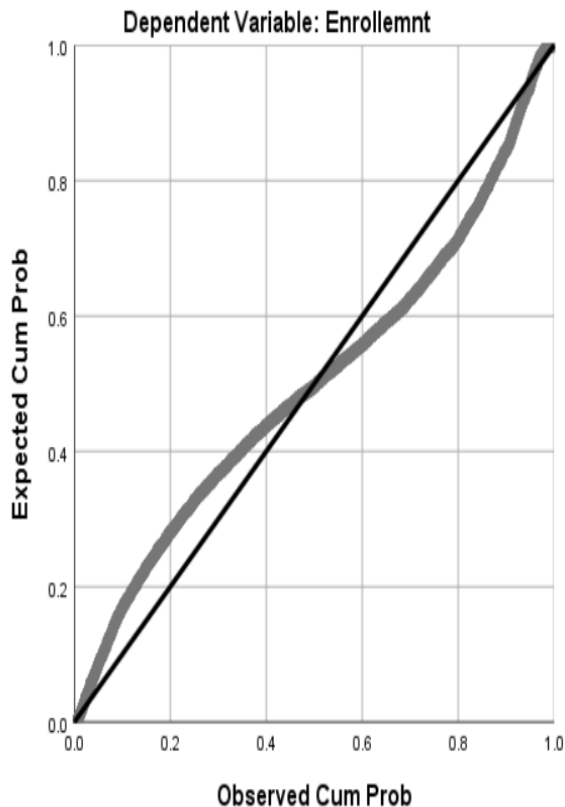
Model Summary[u]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.890[a]	.793	.793	175.286	.793	15921.587	1	4158	.000	
2	.917[b]	.840	.840	154.029	.047	1227.867	1	4157	.000	
3	.929[c]	.863	.863	142.372	.023	709.549	1	4156	.000	
4	.933[d]	.871	.871	138.419	.008	241.762	1	4155	.000	
5	.937[e]	.878	.878	134.597	.007	240.365	1	4154	.000	
6	.938[f]	.880	.880	133.419	.002	74.677	1	4153	.000	
7	.939[g]	.881	.881	132.791	.001	40.342	1	4152	.000	
8	.939[h]	.882	.882	132.344	.001	29.103	1	4151	.000	
9	.940[i]	.883	.883	131.908	.001	28.525	1	4150	.000	
10	.940[j]	.884	.883	131.591	.001	20.975	1	4149	.000	
11	.940[k]	.884	.884	131.306	.001	19.069	1	4148	.000	
12	.941[l]	.885	.884	131.033	.001	18.306	1	4147	.000	
13	.941[m]	.885	.885	130.771	.000	17.637	1	4146	.000	
14	.941[n]	.885	.885	130.582	.000	12.962	1	4145	.000	
15	.941[o]	.886	.886	130.306	.001	18.594	1	4144	.000	
16	.941[p]	.886	.886	130.216	.000	6.720	1	4143	.010	
17	.941[q]	.886	.886	130.131	.000	6.439	1	4142	.011	
18	.942[r]	.886	.886	130.058	.000	5.665	1	4141	.017	
19	.942[s]	.887	.886	130.013	.000	3.866	1	4140	.049	
20	.942[t]	.887	.886	129.961	.000	4.286	1	4139	.038	1.737

Coefficients[a]

Model	B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
Numberofdocumentcameras	6.249	.282	.303	22.179	.000	.890	.326	.116	.147	6.810
AdvanceClassroom	17.859	5.199	.885	3.435	.001	.880	.053	.018	.000	2423.481
Screenresolutionwithhighquality	.202	.012	.180	16.965	.000	.794	.255	.089	.245	4.090
StudentsprovidedEmailAccounts	.122	.007	.127	16.459	.000	.680	.248	.086	.460	2.175
Desktops	.330	.027	.095	12.331	.000	.684	.188	.065	.460	2.173
Numberofmultimediaprojectors	1.703	.209	.093	8.145	.000	.833	.126	.043	.212	4.725
Doyouhaveoneormoregradesinyourschoolthathaveaoneto	-32.120	5.643	-.033	-5.692	.000	.051	-.088	-.030	.796	1.256
NumberofMacOSX10.4orlower	-.534	.184	-.017	-2.910	.004	.067	-.045	-.015	.802	1.247
@1GBorhigher	26.418	4.769	.030	5.539	.000	.263	.086	.029	.908	1.102
NumberofWindowsXPOS	-.503	.111	-.024	-4.510	.000	.113	-.070	-.024	.960	1.041
DeviceswithRAMlessthan1GB	.205	.049	.025	4.165	.000	.052	.065	.022	.752	1.330
Year	-17.404	4.120	-.023	-4.224	.000	.010	-.066	-.022	.957	1.045
ESD	.243	.071	.018	3.430	.001	.010	.053	.018	.965	1.036
NumberofMacOSX10.5orhigher	.236	.049	.027	4.826	.000	.203	.075	.025	.878	1.139
NumberofIOS	-.208	.057	-.075	-3.649	.000	.240	-.057	-.019	.064	15.568
Numberofdigitalmediarecetverse.g AppleTV	-.832	.316	-.016	-2.631	.009	.132	-.041	-.014	.753	1.329
NumberofWindows88.10S	.066	.033	.011	1.982	.048	.122	.031	.010	.899	1.112
ClassroomwithinternetAccess	-12.331	5.190	-.611	-2.376	.018	.880	-.037	-.012	.000	2418.799
NumberofInstructionalTablets	.148	.055	.056	2.700	.007	.275	.042	.014	.064	15.530
NumberofAndroid	-.193	.093	-.014	-2.070	.038	.043	-.032	-.011	.609	1.641

a. Dependent Variable: Enrollemnt

Below ANNOVA model shown below has all independent variable is .000 which means P value is less than .05 indicates statistical significance of the result.

The below coefficient table show that individual contribution of independent variable to

VI. VISUALIZATIONS

Normal P-P Plot of Regression Standardized Residual

Dependent Variable: Enrollemnt

Scatterplot

Dependent Variable: Enrollemnt
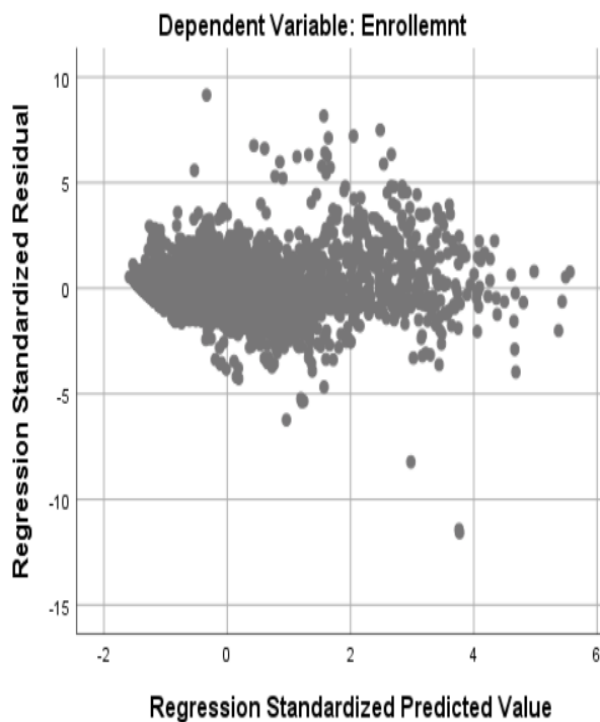
VII. RESULT AND CONCLUSION

It is evident from the above interpretation of model is that from the ANNOVA significance value and coefficient value that it is statistically significant with P value less than 0.5 hence we do not reject null hypothesis. Hence, we can say that tools and technology affect the enrollment of student. So, according to analysis all the schools in particularly district can be benefited from this analysis as here the targeted variable is listed which are going to affect the most in other way it can help to understand the enrollment pattern of student in each school and can appropriate step can be taken from school management to improve the or increase the enrollment. Let us discuss the improvement can be made through the variable if analyzed properly. One of the independent variables is Number of documented cameras which contribute 11 % to dependent variable which is enrollment of student. It can be inferred from this that students and parents are more likely to look safety first over education so, if school improve their security then it is expected the increase in enrollment of student. The model summary says that independent variables have 92.2 % effect on dependent variable which means that the 11 % percent discuss above is likely to occur 92.2 % . In the same manner school administration can look into the other top independent variable which can give them some direction of improvement of their way of approaching the enrollment of student. If we look this in finance perspective, then enrollment and finance go hand in hand. If enrollment will increase finance will also get increase.

This are sample analysis in the same way school administration can try different method to have different view over it.

VIII. REFERENCES

[1] Wladis, C. and Samuels, J., 2016. Do online readiness surveys do what they claim? Validity, reliability, and subsequent student enrollment decisions. *Computers & Education*, *98*, pp.39-56.

[2] Mora, T., Escardíbul, J.O. and Di Pietro, G., 2018. Computers and students' achievement: An analysis of the One Laptop per Child program in Catalonia. *International Journal of Educational Research*, *92*, pp.145-157.

[3] Mensah, D.K.D., 2019. Teachers' Perspective on Implementation of the Double Track Senior High School System in Ghana. *International Journal of Emerging Trends in Social Sciences*, 5(2), pp.47-56.

[4] Investopedia (2019) *Multiple Linear Regression – MLR Definition* [Online] Available at: https://www.investopedia.com/terms/m/mlr.asp [Accessed 23 March 2019].

IX. WORK LOG

Time	Work done
Week 1	Topic research and finalization
Week 4	Finding the datasets
Week 6	Implementation of the method
Week 7	Evaluation of the result and conclusion
Week 8	Report writing