International Journal Of Advance Research, Ideas And Innovations In Technology

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 2)
Available online at: www.ijariit.com

Insightlabs next generation school automation and data analytics

Rahul I.
irahul1997@gmail.com
SRM Institute of Science and Technology, Chennai, Tamil Nadu

Krishna Priya Udaykumar
uk.priyaa@gmail.com
SRM Institute of Science and Technology, Chennai, Tamil Nadu

Leela Sai Kumar
leelasai5@gmail.com
SRM Institute of Science and Technology, Chennai, Tamil Nadu

ABSTRACT

This project is based on a comprehensive study and research of Schools in India, where a school, even after being considered as an organization, perform and consolidate their daily operations such as timetabling, communication, attendance, human resource and accounts using Files and Paperwork. It has been established that there is a growing need that these schools need to go 'digital'. Apart from being digital, it has been found that classrooms generate a lot of data in the form of Academics, Demographic data, Attendance, Classroom Interaction, etc. This data is just given to the students and the parents in the form of a Report Card. This data is considered 'Raw' and useless as it just acts as a notification or information. This project helps Schools visualize and generate Insights through Data Mining and Machine Learning algorithms such as Random forests, etc.

Keywords— *EDA, Data mining, Visualization, Machine learning*

1. INTRODUCTION

School Automation is an integration enterprise software that can empower and streamline the procedures and routine processes of a school beginning from pre-confirmation of the understudies till their Alumni by controlling and improving every one of the exercises in the distinctive offices and connecting those exercises through a centralized distributed cloud powered database medium and furthermore helping the school to turn into the pioneers in the market. The software is a Cloud and Web-based suite of utility and a portion of the sub frameworks are: administration management, student/staff information management system, hr management, marketing management, academic management, student/staff attendance management, library management, inventory, internal assessment management, alumni management, online data dashboard and so on.

Every Classroom of a particular institution is a constant producer and warehouse of data that is being generated. This data can be in the form of student attendance, marks, involvement parameters etc. The current form of recording this data is too inaccurate and insubordinate. The data that is produced is vainly given back to the students and parents through a report card system of Boxing Raw data to Raw legible data. A systemized innovative framework of data to knowledge translation and knowledge management is required and is exactly set to be produced in InsightLabs.

2. VISUALISATION OF ACADEMIC DATA

The data is taken in multiple entries and parameters and is first subjected to an Insightful Visualisation where parents and Students can track student history with ease and grace and also compare performance and various tests. The administration of the school can also have a guided stastistical assistance for thier decision support system. The management can look into aspects of comparison of terms/semesters from the past and make judgements and inferences about the decisions that they have taken. All from a simple visualisation process and framework.

3. DATA ANALYTICS IN EDUCATION - INSIGHT

In order to provide invaluably and needed support to teach- ers and educators, Data Analysis and Machine Learning is basically mining information. The days when educators needed to depend on point by point gradebooks are gone. With Machine Learning, educators approach all their under- study's information in one spot. Notwithstanding conveying a portion of the managerial weight, Data Analytics ad- ditionally enables instructors to improve their exercises by distinguishing where groups of understudies are battling.

A noteworthy advantage of AI is its capacity to foresee understudy execution. By "learning" about every understudy, the innovation can recognize shortcomings and recommends approaches to improve, for example, extra practice tests.

AI can help move far from government sanctioned testing. 'Stop and test' appraisals don't thoroughly assess an understudy's comprehension of a point. The man-made brainpower based appraisal gives steady input to educators, understudies and guardians about how the understudy learns, the help they need and the advancement they are making towards their learning objectives.

AI can likewise review understudies decently by evacuating human predispositions. While evaluating is presently previously being finished by computer based intelligence for numerous decision tests, we are starting to see AI likewise beginning to survey composing with apparatuses like Turnitin and Grammarly.

AI likewise makes it conceivable to modify learning for every understudy in the homeroom. Educators will almost certainly utilize the information to see which understudies need extra help, and the innovation can likewise recommend significant learning devices for every understudy.

Through recognizing shortcomings, AI can sort out sub- stance all the more adequately. For instance, as understudies learn one aptitude, they proceed onward to the following expertise consistently expanding upon the information.

AI, for example, learning investigation, will likewise help improve degrees of consistency. By distinguishing "in danger" understudies, schools can connect with those understudies and get them the assistance they should be effective.

Another way AI will improve training is by gathering understudies and instructors as per their requirements and accessibility. As we find more approaches to use AI in the homeroom, we are discovering more approaches to improve instruction.

4. CLASSIFICATION AND PREDICTIVE ANALYTICS

In order to provide an insight into the data that is produced, we take a data set that can be acquired in a school with the Automation system where data is entered by the teachers and the parents as well as captured from their tracked behavioral analysis.

The above information is collected and gathered by using the existing data that a student produces on a daily, weekly or annual basis and a student action tracker apparatus, which is also called an x-API. The xAPI is a part of the preparation and learning process (TLA) that completely empowers to screen learning advancement and student's activities. The dataset that is used for this project comprises of 480 understudy records and 16 highlights. The high- lights are grouped into three noteworthy classes:

(a) Statistic highlights, for example, sexual orientation and nationality.
(b) Scholarly foundation highlights, for example, instructive stage, grade Level and area.
(c) Conduct highlights, for example, raised a hand on class, opening assets, noting a study by guardians, and school fulfillment. This helps in assessing the behaviour of a student.

The dataset comprises of 305 guys and 175 females. The understudies originate from various sources, for example, 179 understudies are from Kuwait, 172 understudies are from Jordan, 28 understudies from Palestine, 22 understudies are from Iraq, 17 understudies from Lebanon, 12 understudies from Tunis, 11 understudies from Saudi Arabia, 9 understudies from Egypt, 7 understudies from Syria, 6 understudies from USA, Iran and Libya, 4 understudies from Morocco and one understudy from Venezuela.

The data set is gathered through two instructive semesters: 245 understudy records are gathered amid the main semester and 235 understudy records are gathered amid the second semester. The informational index incorporates likewise the school participation highlight, for example, the understudies are ordered into two classifications dependent on their non-attendance days: 191 understudies surpass 7 nonappearance days and 289 understudies their nonappearance days under 7.

```
gender                    0
NationalITy               0
PlaceofBirth              0
StageID                   0
GradeID                   0
SectionID                 0
Topic                     0
Semester                  0
Relation                  0
raisedhands               0
VisITedResources          0
AnnouncementsView         0
Discussion                0
ParentAnsweringSurvey     0
ParentschoolSatisfaction  0
StudentAbsenceDays        0
Class                     0
dtype: int64
```

Fig. 1: The parameters for the data is produced. The data type for the entries is subjected to integer - 64 bytes

Characteristics

(a) Sex - understudy's sexual orientation (ostensible: 'Male Student' or 'Female Student')
(b) Nationality- comprises the understudy's nationality (ostensible:' Kuwaiti',' Lebanon',' Egypt',' Saudi Ara- bia',' USA',' Jordan',' Venezuela',' Iran',' Tunis',' Morocco',' Syria',' Palestine',' Iraq',' Lybia')
(c) Spot of birth-understudy's Place of birth (ostensible:' Kuwait',' Lebanon',' Egypt',' Saudi Arabian',' USA',' Jordanian',' Venezuelanian',' Irani',' Tunisian',' Moroccan',' Syrian',' Palestinian',' Iraqi',' Lybian')
(d) Instructive Stages-instructivedimension understudy has a place (ostensible: 'lower- level', 'Middle School', 'High School')
(e) Evaluation Dimensions grade understudy has a place (ostensible: 'Gr01', 'Gr:02', 'Gr:03', 'Gr:04', 'Gr:05', 'Gr:06', 'Gr:-07', 'Gr:-08', 'Gr:-09', 'Gr:-10', 'Gr:-11', 'Gr:-12')
(f) Area ID-study hall understudy has a place (nominal: 'A','B','C')
(g) Theme course subject (ostensible:' English',' Spanish', 'French',' 3rdLanguade',' Information Technology',' Math',' Science', 'Science', 'Science',' History',' Religious studies',' Geography')
(h) Semester-school year semester (ostensible:' First' or ' Second')
(i) Parent in charge of understudy (nominal: 'Maternal (Mother)','Paternal '(Father)
(j) Raised hand-how often the understudy raises his/her hand on study hall (numeric:0-100)

(k) 11-Visited assets how often the understudy visits a course content(numeric:0-100)
(l) 12 Survey declarations how often the understudy checks the new announcements(numeric:0-100)
(m) 13 Talk gatherings how often the understudy take an interest in discourse gatherings (numeric:0-100)
(n) 14 Parent Noting Study parent addressed the studies which are given from school or not (nominal:'Yes','No')
(o) 15 Parent School Fulfillment the Level of parent fulfillment from school(nominal:'Yes','No')
(p) 16 Understudy Nonattendance Days-the quantity of nonappearance days for every understudy (ostensible: over 7, under-7)

We start visualizing raw data of a set of students, subject wise and grade wise. We start categorizing failures and no. of failures mapped to their respective subjects.

```
```

Code Snippet 1.

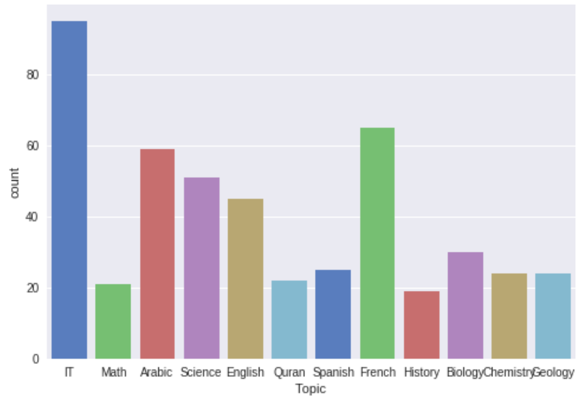

Fig. 2

The understudies are characterized into three numerical interims dependent on their complete evaluation/mark:
L: interim incorporates values from 0 to 69,
M: interim incorporates values from 70 to 89,
H: interim incorporates values from 90-100

In relational plots, we perceived how to utilize distinctive visual portrayals to demonstrate the connection between different factors in a dataset. In the precedents, we concen- trated on situations where the fundamental relationship was between two numerical factors. On the off chance that one of the principle factors is "clear cut" (separated into discrete gatherings) it might be useful to utilize a progressively specific way to deal with representation.

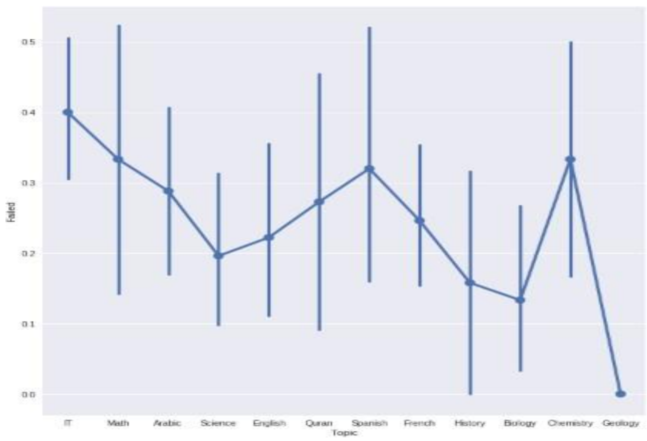

Fig. 3: A seaborn scatter plot giving a relational comparison of the failures with respect to subjects

Topic	Arabic	Biology	Chemistry	English	French	Geology	History	IT	Math	Quran	Science	Spanish
Class												
H	19	16	10	17	20	6	4	15	6	8	16	5
L	17	4	8	10	16	0	3	38	7	6	10	8
M	23	10	6	18	29	18	12	42	8	8	25	12

Fig. 4: It shows in this table that not one member had failed Geology while the respective students in Information Technology, Math, and Chemistry had the most significant chance of failing in the subject

During the entire sample analysis, we use python to train the learning model and also visualize the entire parameter database.

```
sns.countplot(x='Class',data=df,palette='
    BuPu"')plt.show()
```

Code Snippet 2. We make a comprehensive bar representation of the above data.

As we now have consolidated a count for the Medium, High and Lower Class, We can now take another very important variable with Parent School Satisfaction and com- pute it with the Class. This will now help us understand the role that parents play with School Satisfaction and how that translates to good academic results.

```
sas.countplot(x='
    ParentschoolsSatisfaction',data = df,
     hue='Class',palette='bright')plt.
    show()
```

Code Snippet 3. Visualizing the Parent School Satisfaction parameter with class.

So from conferring with the above results, we can now come to one particular conclusion. Parents who are actually satisfied with their respective schools generally and are proved to have children who perform better in their academics. We had also concurred that the least satisfied parents had their children do relatively worse in their academics.

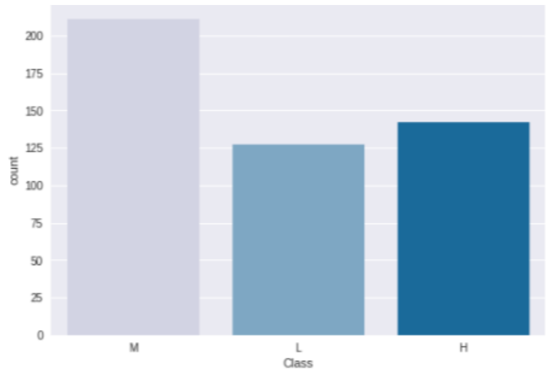

Fig. 5: Showing the total grade level count throughout all students

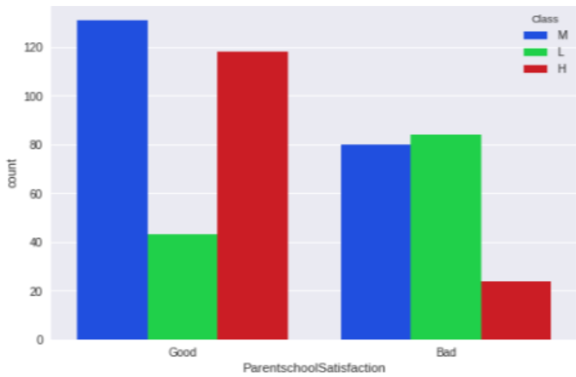

Fig. 6: Parent school satisfaction computed with now the count of Students with their academic class

We now go to the relation between the responsible parent sex vs. the failure rate of the dataset.

```
sas.factorplot('Relations','Failure',data
    =df)
```

Code Snippet 4.

The students who had their Mothers responsible for them in school had the biggest chance of performing well in school. Such an Insight just by the right visualization can go a long way in seeing things in perspective and changing for the better.

We can now see that there is a higher percentage of the male population failing with comparison to the female popula- tion. This can be condoned due to the higher majority of the male population. If not, the school's policy in treating the male students should be taken a look into. This kind of visualization process can help the administration of a school do wonder with amping up academic success.

Let us now compare Raised Hands with the Academic Success of the students. Raised hands can be a helpful quantitative parameter for studying and analyzing the involvement of students in the school.

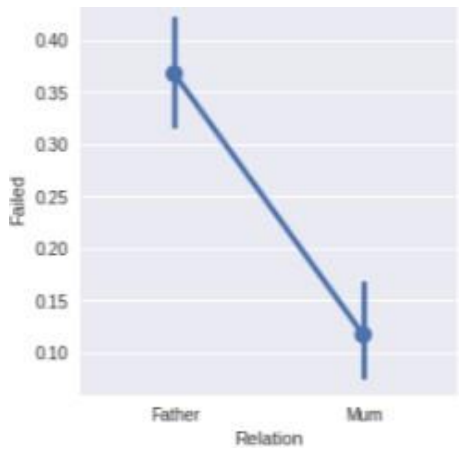

Fig. 7: We now see a representation between the failures and the students who had mothers or fathers as their first responsible person in school

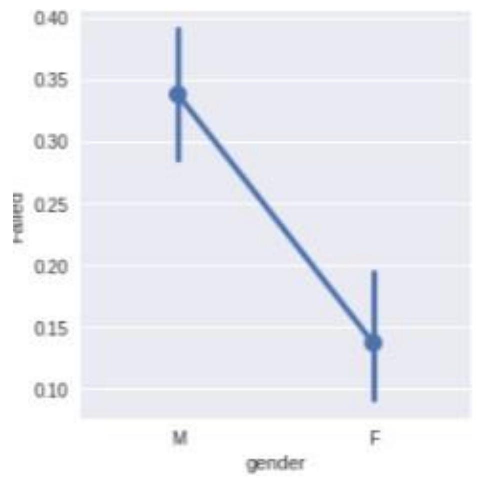

Fig. 8. Here is a comparison of the students' gender with the number of failures

```
Raised_hands = sns.boxplot(x="Class", y="
    raisedhands", data=df)Raised_hands=
    sns.swarmplot(x="Class", y="
    raisedhands", data=df, color=".13")
    plt.show()
```

Code Snippet 5. Python Code for displaying the relation, we use a swarm plot to compute the exact placement vs. the academic success with respect to raised hands

```
df.groupby('Topic').median()
```

Code Snippet 6. We now take the median of the parameters to perceive further calculations

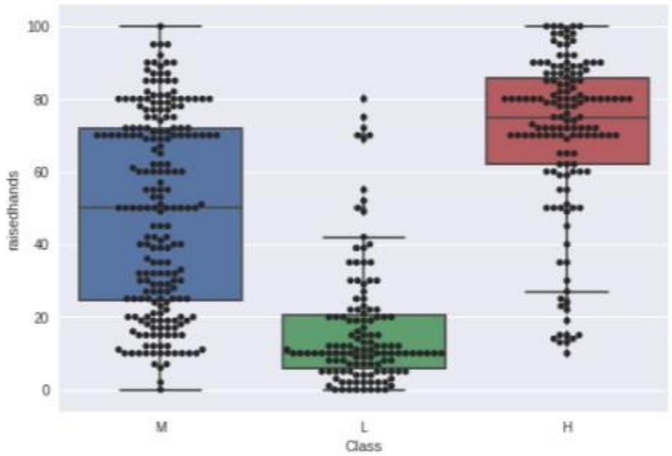

Fig. 9: This swarm plot and box plot shows a clear comparison between raised hands and Class MLH.

	raisedhands	VisITedResources	AnnouncementsView	Discussion	Failed
Topic					
Arabic	32.0	65.0	41.0	38.0	0.0
Biology	78.5	88.5	54.0	47.0	0.0
Chemistry	79.0	84.5	47.0	30.5	0.0
English	55.0	50.0	33.0	36.0	0.0
French	35.0	80.0	23.0	21.0	0.0
Geology	80.0	82.0	68.5	60.5	0.0
History	69.0	84.0	72.0	65.0	0.0
IT	20.0	25.0	10.0	40.0	0.0
Math	28.0	15.0	19.0	40.0	0.0
Quran	65.0	75.0	50.0	45.0	0.0
Science	62.0	64.0	58.0	66.0	0.0
Spanish	27.0	51.0	40.0	20.0	0.0

Fig. 10

5. LEARNING THROUGH A SUPPORT VECTOR MACHINE

The most comprehensive and material Machine learning calculation for our concern is a Linear SVC where we must begin by training perceptrons.

The goal of a Linear SVC (Support Vector Classifier) is to fit the information given, restoring a "best fit" hyperplane that isolates, or arranges, your information. From that point, in the wake of getting the hyperplane, you would then be able to nourish a few highlights to the classifier to perceive what the "anticipated" class is. This makes this particular calculation somewhat appropriate for our utilization.

Putting the classification as our first priority, we now start by training a perceptron.

```
df['TotalQ'] = df['Class']df['TotalQ'].
    loc[df.TotalQ == 'Low-Level'] = 0.0df
    ['TotalQ'].loc[df.TotalQ == 'Middle-
    Level'] = 1.0df['TotalQ'].loc[df.
    TotalQ == 'High-Level'] = 2.0
    continuous_subset = df.ix[:,9:13]
```

```
    continuous_subset['gender'] = np.
    where(df['gender']=='M',1,0)
    continuous_subset['Parent'] = np.
    where(df['Relation']=='Father',1,0)X
    = np.array(continuous_subset).astype
    ('float64')y = np.array(df['TotalQ'])
    X.shape
from sklearn.cross_validation import
    train_test_splitfrom sklearn.
    preprocessing import
    StandardScalerX_train, X_test,
    y_train, y_test = train_test_split(
            X, y, test_size=0.3,
    random_state=0)sc = StandardScaler()
    sc.fit(X_train)X_train_std = sc.
    transform(X_train)X_test_std = sc.
    transform(X_test)
from sklearn.linear_model import
    Perceptronppn = Perceptron(n_iter=40,
     eta0=0.1, random_state=0)ppn.fit(
    X_train_std, y_train)y_pred = ppn.
    predict(X_test_std)print('
    Misclassified samples: %d' % (y_test
    != y_pred).sum())
from sklearn.linear_model import
    Perceptronppn = Perceptron(n_iter=40,
     eta0=0.1, random_state=0)ppn.fit(
    X_train_std, y_train)y_pred = ppn.
    predict(X_test_std)print('
    Misclassified samples: %d' % (y_test
    != y_pred).sum())
```

Code Snippet 7. Training the SVM perceptron and also capturing misclassified samples

Fig. 11: Misclassified Samples have been identified

```
from sklearn.metrics import
    accuracy_scoreprint('Accuracy: %.2f'
    % accuracy_score(y_test, y_pred))
from sklearn.metrics import
    classification_reportprint(
    classification_report(y_test, y_pred)
    )
```

Code Snippet 8. Directing accuracy

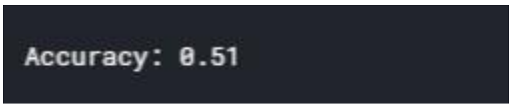

Fig. 12: Accuracy is above satisfactory for a small data set that is being practiced upon

From here we can see a huge difference that actually attending school makes with respect to the correlation between Absence and Academic Success. Here shows a clear verified factor that absolutely defines how important attendance is.

Table 1: Measuring Accuracy and Verification

	Precision	Recall	F1-score	Support
H	0.46	0.72	0.56	39
L	0.6	0.53	0.56	34
M	0.51	0.38	0.44	71
Avg/Total 0.52		0.51	0.50	144

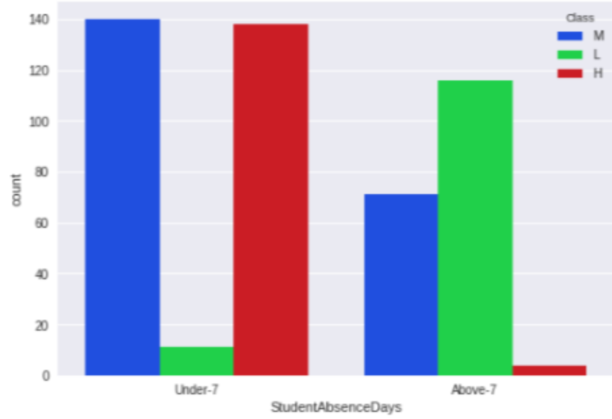

Fig. 13. A computation visualization of LMH with the Absence and Count

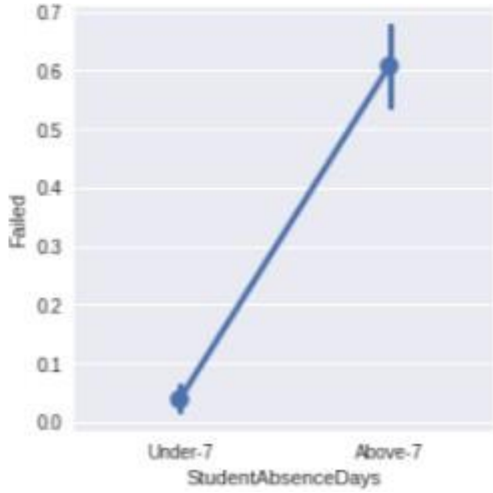

Fig. 14. SeaBorn plot comparing Attendance and Failure

The greatest visual pattern can be found in how much of the attendance the understudy was missing. Over 90% of the understudies who did inadequately were missing class in excess of multiple times, while practically none of the understudies who did well were missing in excess of multiple times.

We will make a variable for this class, and incorporate it in our model.

In spite of the fact that parent fulfillment and involvement demonstrated an immense example as for how well an understudy student had actually performed in the class, there is no data on whether the review was taken after evaluations were posted, and moreover the data set does not give any data about the understudy's study hall conduct so it was ignored and condoned in the study.

6. CONCLUSION

From what we have referenced and done in the above exploratory visualization and data analysis. We can actually conclude many possible factual interpretations.

- Girls are to have performed well in academics in comparison to that of boys
- In the case of female students, it is the mothers that take an interest and involvement in their education than the fathers.
- Girls had also fared better than boys in terms of attendance

No obvious sex or gender bias/ inclination with regards to subject/point decisions, we can't infer that young ladies performed better since they maybe took less specialized subjects

Sexual orientation uniqueness holds even at a nation level. And a sharp exploratory data analysis of the data proves that the country's development politically, culturally and socially proves to play a huge factor in the schooling and academic success.

EDA or Exploratory Data Analysis is crucial in playing a strong role in influencing a schools decision making and also redefining government backed educational policies. An all-round School Automation System supported by sharp visualization perception and EDA should be a must and should be key in the countries development.

7. REFERENCES

[1] Khlopotov M. V. Models and algorithms of educa- tional data mining for decision support:
[2] University of Information Technologies, Mechanics and Optics. – Saint-Petersburg, 2014. –127 p.
[3] Stasyshin V.M., Avrunev O.E., Afonina E.V., Lyax K.N. Informacionnaja Sistema experience and current status]. Otkrytoe i distan- cionnoe obrazovanie
[4] The Tenth International Conference on Educational Data Mining EDM 2017.
[5] JEDM - Journal of Educational Data Mining.
[6] Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. International Journal of Database Theory and Application, 9(8), 119-136.Amrieh, E. A., Hamtini, T., & Aljarah, I. (2015, November).
[7] Preprocessing and analyzing educational data set using X-API for improving student's performance. In Applied Electrical Engineering and Computing Technologies (AEECT), 2015 IEEE Jordan Conference on (pp. 1-5). IEEE.