

Performance Analysis of Students from Private and Government Schools of Kathmandu Valley

Sampada Dhakal ^a, Arun Kumar Timalina ^b

^{a, b} Department of Electronics and Computer Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal

Corresponding Email: ^a sampadadhakal05@gmail.com, ^b t.arun@ioe.edu.np

Abstract

Private and public schools have existed in Nepal for a long time. With the difference in school type, the performance of students has been observed to be different as well. To prove this well accepted fact on the mathematical basis, analysis has been done in this project using several tools provided by statistics and data analysis. To estimate the probability density of obtained marks of students, kernel density estimation has been used. The marks distribution showed students in private schools obtain comparatively higher marks. Z-score test showed gender did not affect the marks of students. Linear regression technique has been implied to find out if the final term marks can be predicted on the basis of earlier exams. The results showed the prediction of final term marks can be done with greater accuracy for private schools. Similarly, correlation coefficients have been calculated to find whether there exists a relationship between marks obtained at various subjects which showed students tend to obtain less marks at Mathematics than any other subjects.

Keywords

data analysis, significance test, regression, kernel density estimation

1. Introduction

Nepal Law Commission defines government schools as those schools that have obtained approval or permission for operation and receiving regular grants from the Government of Nepal and private schools as those schools that have obtained approval or permission for operation on condition that they are not receiving regular grants from Government of Nepal.[1] According to Nepal Law Commission, primary education means education given from class one to class five. Lower Secondary Education means the education given from class six to class eight. And, secondary education means education to be given from class nine to class ten. [1]

Government and private schools have existed in Nepal for a long time. It is a well known fact that students in private schools perform way better than students at government schools. Though private school has been the dream of almost all the students and parents, its expensive nature has been the curse to get admitted in one of them. Despite the government's efforts to provide free education for every child, for education being the most important of all the basic needs, the country still lies far behind in its goal. To get a quality

education is termed as being lucky although it is one of the needs every child should be blessed with [2].

Nationwide study has been done regarding the performance of students appearing for grade 10 examinations. [3]. Various data analytics tools can be used to improve the results. Educational Data Mining (EDM) is an emerging field, exploring data in educational context by applying different data mining tools and techniques[4]. Prediction, clustering, relationship mining, distillation of data for human judgement, discovery with models are the five different steps involved in Educational Data Mining [5]. In [6], authors propose features that have a great influence on student's academic achievement by using classifiers such as Naive Bayes, Decision Tree and Artificial Neural Network. Using various tools and techniques, the authors propose indicators for good or poor performance of students of undergraduate students [7]. Various association mining algorithms and clustering techniques have been used in [8] to extract insights from 11 attributes present in 666 instances from the Common Entrance Examination of India.

Particularly there are different reasons for the low

performance of the students in government school and the reasons for students of private school securing higher marks[2].

2. Theoretical Background

2.1 Kernel Density Estimation

Kernel density estimation is a technique for estimation of probability density function that is a must-have enabling the user to better analyse the studied probability distribution than when using a traditional histogram. Unlike the histogram, the kernel technique produces smooth estimate of the pdf, uses all sample points' locations and more convincingly suggest multimodality. In its two-dimensional applications, kernel estimation is even better as the 2D histogram requires additionally to define the orientation of 2D bins. Two concepts play fundamental role in kernel estimation: kernel function shape and coefficient of smoothness, of which the latter is crucial to the method[9].

2.2 Linear Regression Model

Linear Regression Model is used in machine learning tasks to predict the outcome when the required independent variables are provided. It assumes the linear relationship between feature and target variables. Here, the feature variables refer to independent variables whereas target variables refer to dependent variables. In short, if the dataset can satisfy the condition, $Ax = b$ with low error, the model trained with varieties of data can accurately predict the dependent variable when new independent variables are provided.

2.3 Pearson's Correlation Coefficient

Correlation is the method to examine the relationship between two continuous and quantitative variables. Out of different types of correlation coefficient, Pearson's correlation coefficients denote the strength of relationship between variables in a dataset. The coefficients value range from -1 to 1 where -1 denotes strong negative relationship, 1 denotes strong positive relationship and 0 denotes no relationship.

2.4 Significance Test

Z-score is used in statistics to check the significance which helps to decide whether or not to reject the null

hypothesis. In general, the null and alternate hypothesis are chosen, and in reference to the provided data, z-score is calculated. According to the calculated z-score, p-value is found out, where p-value refers to the probability. The particular p-value is chosen as the critical value, below which the null hypothesis is discarded.

3. Methodology



Figure 1: Block diagram of overall process

The overall block diagram of the methodologies is depicted by the figure 1. The explanation of the listed steps are shown in following subsections:

3.1 Data Collection

Student's data from classes 5-9 were collected from 2 private schools (Canvas Basanta Ritu School and Sarswoti Boarding School) and 2 government schools (Nepal Yubak Secondary School and Shree Shanti Bidhya Griha). The collected data had 761 data entries from government school and 1089 data entries from private school. The marks obtained at the important subjects like English, Nepali, Mathematics, Science and Social were obtained, whereas the gender of the students were also acquired.

3.2 Data Cleaning

For the data cleaning, following steps were applied. First of all, the provided handwritten data was digitized. The entered data was too messy, after rejecting the null values, the new count was obtained. For private school, there was no absentees in the examination, so there was no change in the final output. Whereas there was a vast reduction in the number of students after discarding the null values in government schools i.e from 761 to 599. After this process, data was cleaned according to the requirement of the particular research questions.

The distribution of total number of students at private and government schools at different grades can be seen from table 1.

Table 1: Total number of students at different classes after data cleaning

School Type	Private	Government
Five	261	105
Six	195	120
Seven	219	126
Eight	225	147
Nine	189	102
Total	1089	600

3.3 Descriptive/ Predictive Analysis

According to the objectives of the project, different techniques were applied to conclude the result. To visualize the distribution of marks according to classes and school type, kernel density estimation plot was plotted. To understand the significance of gender in the marks, z test was performed. To predict the third terminal marks on the basis of first and second terminal, linear regression model was used. To understand the relationship between class marks at different subjects, Pearson’s correlation coefficient was calculated.

4. Results and Analysis

4.1 Distribution of marks of students according to class and subjects

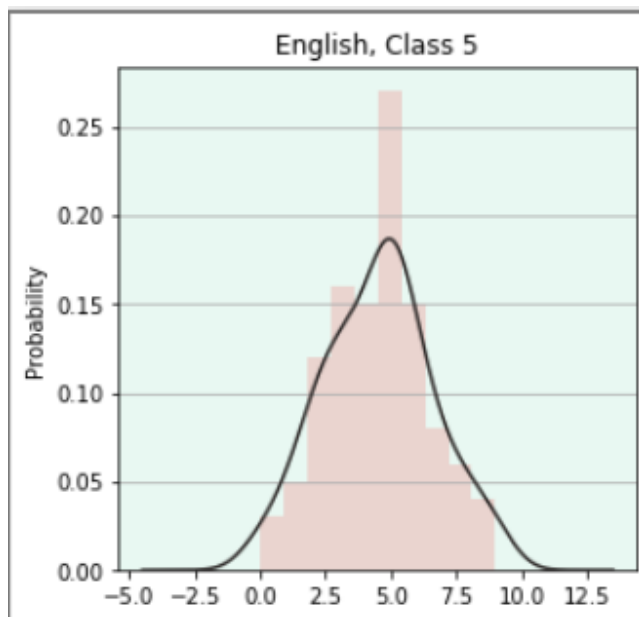


Figure 2: kde of marks obtained by students at English from grade 5 of Government school

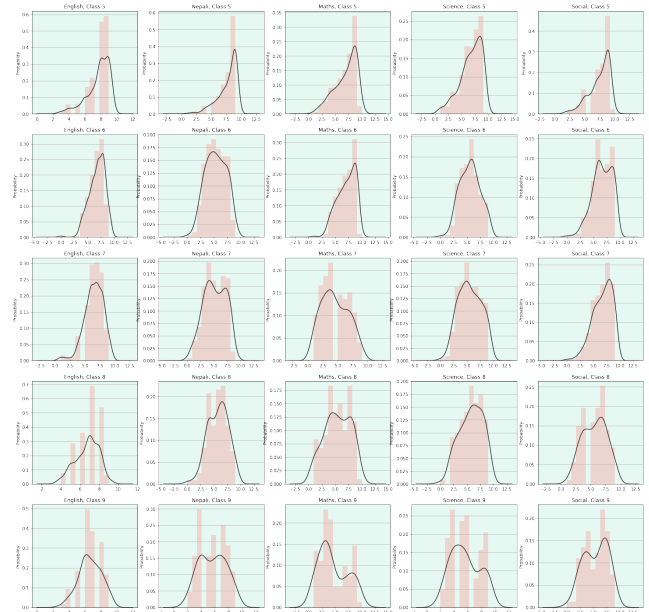


Figure 3: kde of marks according to class and subjects of private school

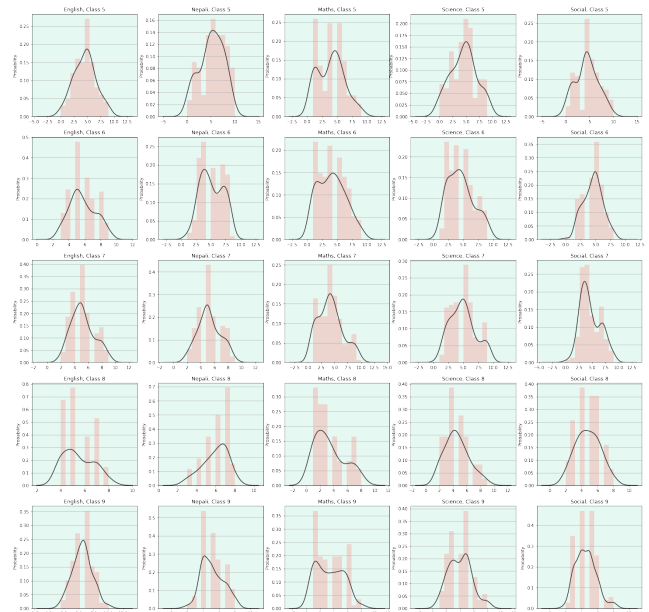


Figure 4: kde of marks according to class and subjects of government school

Figure 2 shows the probability distribution of marks obtained at English by the students of grade 5 from government school. The comprehensive probability distributions are shown by the figures 3 and 4.

From figure 3, it can be seen that for private schools, the kde plot is negatively skewed, implying that there are more students securing higher marks. The marks obtained by students at smaller classes is greater whereas the marks start to gradually decrease when

the students level up to the next class. The decrease in marks for English, Science and Social is gradual whereas for Mathematics and Nepali, marks seem to reduce drastically.

From figure 4, it can be observed for government school, that the kde plot is more or less normally distributed whereas the skewness being sometimes positive and sometimes negative by a small fraction. According to the class, there seems to be some fluctuation in marks secured at English, Nepali, Science and Social and the overall pattern can be seen as the majority of students securing the marks around 40-70. But for the subject Mathematics, marks seem to lower according to leveling up the classes.

4.2 To check whether gender affects the overall performance of the students.

To find out if gender affects the performance of primary and secondary level students, Z-test was performed on various sets of data. Setting p-value as 0.05, the results are obtained as follows:

School Type	Private	Government
English	3 S, 2 NS	1S, 4 NS
Nepali	4 S, 1 NS	1 S, 4 NS
Mathematics	1 S, 4 NS	1 S, 4 NS
Science	2 S, 3 NS	1 S, 4 NS
Social	2 S, 3 NS	1 S, 4 NS

Table 2: Comparison between significance of gender in obtained marks of private and government schools

Here, S is Significant and NS is Not Significant. It can be seen from table 2, that for private school, the significance of gender in marks seem fluctuating, whereas for government school, though there are some cases of significance, majority of the result shows that gender has no role in the obtained marks. It can be concluded that gender does not play a major role in obtained marks of students in primary and secondary levels.

4.3 Prediction of third terminal marks on the basis of first and second terminal marks.

To predict the marks of third terminal on the basis of first and second terminal, linear regression model was trained with 80% of the data. The test was done with 20% of the data and it shows the following result.

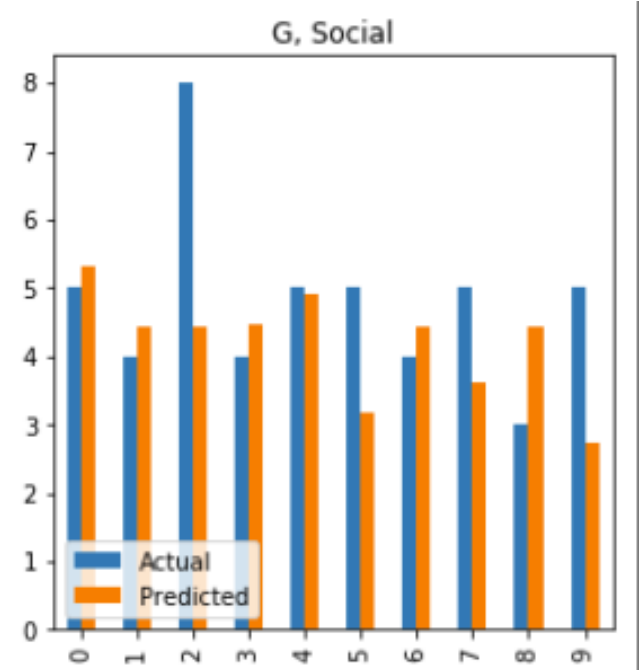


Figure 5: Prediction of third terminal marks on the basis of first and second terminal of social studies of students from government school

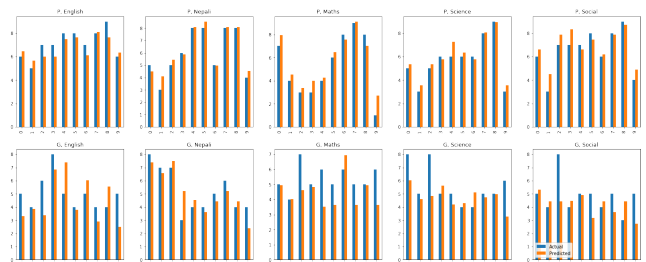


Figure 6: Prediction of third terminal marks on the basis of first and second terminal

Subjects	Private	Government
English	0.78	1.47
Nepali	0.62	1.35
Mathematics	0.67	1.91
Science	0.64	1.68
Social	0.76	1.46

Table 3: RMSE of different subjects

Here, RMSE means Root Mean Squared Error

In reference to figure 6, here, P means students of Private School and G means students of Government School. The subjects following P and G means marks obtained by students from a particular type of school i.e government or private on different subjects. The

linear regression model with the equation $Ax = b$ is used to predict the marks. As the label is shown at the right bottom of the chart, there are two bars i.e actual marks and predicted marks. It can be observed that the prediction of third terminal marks on the basis of first and second terminal marks can be done with greater accuracy for private school, whereas the accuracy is comparatively smaller for government school. The details of the figure can be observed from figure 5. It can also be illustrated from the table 3 that though the marks can be predicted after fitting into a linear regression model, the accuracy varies for government and private schools.

4.4 Correlation between marks obtained at different subjects.

To find out whether the relationship exists between marks obtained at one subject to the other, first of all, correlation coefficient was calculated among all the subjects and secondly, the correlation coefficient with every subjects to the overall marks was calculated.

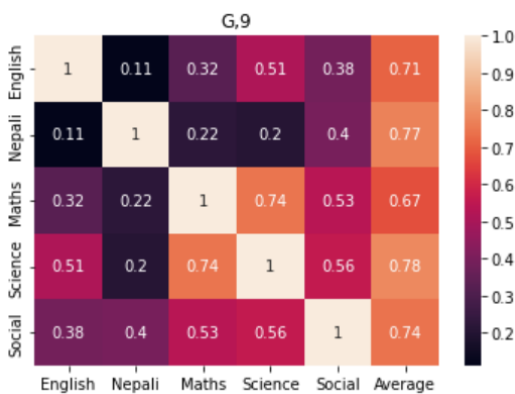


Figure 7: Correlation between marks at different subjects of grade 9 students at government school

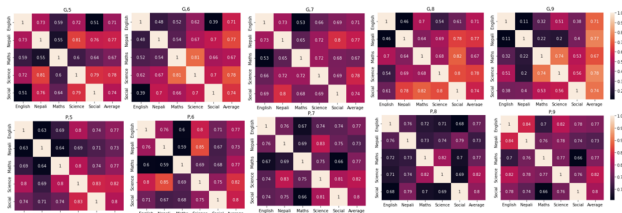


Figure 8: Correlation between different subjects

In figures 7 and 8, G and P means Government and Private schools respectively and the number 5, 6, 7, 8 and 9 followed represents the class. In figure 7, the relationship between marks obtained at different subjects by students of grade 9 from government

school is shown. Figure 8 shows the comprehensive strength of relationship between marks obtained from one subject to another at different grades. The small correlation value with Mathematics shows that the pattern of marks obtained by student is different than marks obtained at other subjects. This can also be explained with reference to figures 3 and 4 by the positively skewed plots at Mathematics.

5. Conclusion

From the analysis done in the research, it can be seen that there are more students dropping exams and classes in government school whereas this trend is not observed for the students of private school. Kernel density estimation showed the difference in performance of students in various grades and z test showed gender does not play any role at all. Linear regression model could predict the final terminal marks of students with higher accuracy for private school than for government school. The correlation coefficients showed students tend to obtain lesser marks in Mathematics than any other subjects in both private and public schools.

While few features are taken into consideration while carrying out the work, there are other factors affecting the performance of students such as regularity of teachers and students, economic background of the students, family education etc. These variables can also be accounted while deriving the results.

Acknowledgments

We would like to give special thanks to Canvas Basanta Ritu School, Sarswoti Boarding School, Nepal Yubak Secondary School and Shree Shanti Bidhya Griha for providing the data needed to conduct this research. We would like to thank Mr. Krishna Prasad Dhakal for helping us collect the datasets from various schools.

References

- [1] Nepal Law Commission. Nepal law commission, 1971.
- [2] Amit Koirala. Debate on public and private schools in nepal. *International Journal of Social Sciences and Management*, 2:3–8, 01 2015.
- [3] Ministry of Education and Sports. Study on student performance in slc, 2006.
- [4] Rajni Jindal and Malaya Dutta Borah. A survey on educational data mining and research trends. *International Journal of Database Management Systems*, 5(3):53, 2013.

- [5] Abdulmohsen Algarni. Data mining in education. *International Journal of Advanced Computer Science and Applications*, 7(6):456–461, 2016.
- [6] B Sana, Isma Farrah Siddiqui, and Qasim Ali Arain. Analyzing students' academic performance through educational data mining. 2019.
- [7] Raheela Asif, Agathe Merceron, Syed Abbas Ali, and Najmi Ghani Haider. Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113:177–194, 2017.
- [8] Sadiq Hussain, Rasha Atallah, Amirrudin Kamsin, and Jiten Hazarika. Classification, clustering and association rule mining in educational datasets using data mining tools: A case study. In *Computer Science On-line Conference*, pages 196–211. Springer, 2018.
- [9] Stanislaw Weglarczyk. Kernel density estimation and its application. *ITM Web of Conferences*, 23:00037, 01 2018.