

# Framework for Performance Measurement of Public and Private Sector Higher Education Institutions in Pakistan using Machine Learning Algorithms

Muhammad Zulqarnain Siddiqui  
School of Science and Engineering

Malaysia University of Science and Technology, Malaysia

Dr. Sellappan Palaniappan  
School of Science and Engineering

Malaysia University of Science and Technology, Malaysia

**Abstract:-** This study proposes the mechanism to collect real time data relating to education sector into a centralize repository and then to process specific portion of collected data using machine learning algorithms to compare performance of public and private sector higher education systems in Pakistan. Data of selected batch of students from colleges and universities are interlinked through a central repository that organizes and processes data to monitor the overall performance in terms of student retention gives trends of admissions and drop outs. A framework to that will facilitate the stakeholders including Government, Students and institutions to overview the progress at any level and can be sliced to any institution as well.

**Keywords:-** Machine Learning, Repository, Distributed Systems, Data Sharing, Algorithms.

## I. INTRODUCTION

Pakistan is one of the developing countries around the world having total population of over 200 million, having literacy rate of 58%. As indicated by economic survey of Pakistan in 2018 one of the major concerns in the progress is the rate of literacy that needs to be improved on priority basis in order to transform the nation.

Current educational structure of Pakistan is divided into different levels starting from primary level schooling, secondary schools, colleges and universities. Citizens who have attended primary level education are included in 58% literacy rate. It has been observed over the years that most the population does not pursue their studies after schooling. Hence there is drastic change of numbers in terms of registrations of colleges after schooling. There are many students who drop outs in colleges and universities after taking admissions there.

Elementary education starts from pre-primary level schooling, followed by primary schooling (having classes from classes 1-5), middle schools (classes 6-8), high schools (classes 9 and 10) and higher secondary schools termed as colleges. After high schools (class 10) student is titled as a school pass out. Most of the school pass outs from private sector institutions continues their higher secondary schooling whereas the ratio is comparatively low in public sector schools. Following figure 1 represents the structure of basic schooling in Pakistan:

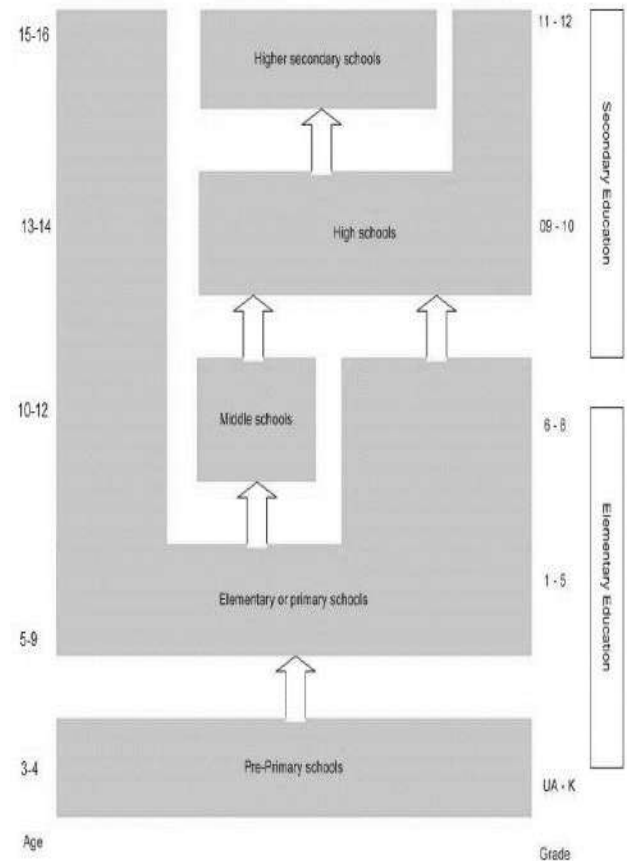


Fig 1:- Basic Education System <sup>[1]</sup>

There is no central automated mechanism to collect real-time data of all educational institutions from where this analysis can be carried out. A report published by Educational ministry of Pakistan in year 2018 gives the analysis on the facts collected in year 2005 and 2006. Hence there is a strong need of integration of educational data to central repository in order to monitor the progress of institutions and individuals, to find out the pattern of pass out students and drop out ratio and to find the trend of students intending to pursue studies after schooling. This data source can be beneficial to all stakeholders including students, institutions and government to monitor and plan for the progress of educational sector.

There are many reasons for not having higher rate of literacy, including ghost attendance, students bunking classes, cheating in examinations etc. are highly witnessed

in schools and colleges in different cities of Pakistan. In order to improve the current situation of the education system, a centralized automated academic source is needed. This academic portal that can keep the records of all these educational institutes (private and public), and by processing this data, we can forecast various critical points regarding improvement in education sector.

Proposed system will bring all the academic data of a student enrolled in any level of education, on a single platform, where all the stakeholders will have access to it remotely. It will facilitate parents, teachers, and higher education organizations like HEC, different educational board, government and foreign institution to analyze the current educational trends and have sufficient pre-plans for improving the quality of education. This project will be beneficial for students to take a decision on their academic itineraries through data mining as well as it would be beneficial for the parents and government as well.

It has been observed that there are enormous problems in the education system of Pakistan especially in the Government sector institutions. Usually, in government schools or colleges there is a huge number of absenteeism of teachers and students on regular basis but managed smoothly on papers for the record. The attendance of the teacher or students is adequate, though they were not present in school or college to teach or study. According to the Education Minister of a province, there were around 900 ghost schools with no records of 15000 ghost teachers in entire state. Due to the malingering or lack of interest of teachers, students also suffer. Teacher rather than paying attention to the students or encouraging students, they doomed the student's knowledge. The education sector did not give importance to the skills or knowledge of teachers while hiring them. Lack of skills and training are also factors due to which teachers do not provide proper guidance to students for their future. Proposed system sets up to perform as a liaison between students, teachers and the whole academia. Focusing on selected number of institutions and students this prototype has been used for the analysis of real time data.

## II. LITERATURE REVIEW

Implementation of informatics in education sector of Turkey started with the initiative of Government to commence a special program "Specialized Commission on Computer Education at Secondary Schools". In the initial phase of the program purchase and procurement of hardware, development and deployment of customized software and training of teachers and staff of focused institutions were carried out. In the framework of "Project on the Development of National Education" supported by World Bank, projects on "Computer Piloting Schools" which covered 53 schools, and "Computer Laboratory Schools" which covered 182 schools were conducted with the purpose of the expansion of use of computers, and computer assisted education.<sup>[3]</sup>

Different training and awareness sessions to make the stakeholders including teachers and students use technology in routine academic activities. Teachers specially were targeted to inculcate technology in their lectures and assignments. This will help to maintain the repository of all data of assignments, quizzes and examination<sup>[4]</sup>.

Institutions and schools (along with their computer laboratories) that are linked and managed by Ministry has been provided with hi-speed internet services in order to maintain the infrastructure enabling smooth IT services for distance education. Students that registers with one the schools or institutions were given the access to e-books for selected courses<sup>[7]</sup>.

Project for automation was divided into four phases. In the initial three phases around 20000 institutions were equipped with Asymmetric digital subscriber line (ASDL) internet access and this was expanded to other institutions in later years. Batch of twenty thousand teachers were selected in the initial phase to be educated trained to efficiently use the system supported by Ministry. In the next phase batch of 20000 teachers each year were focused for the same training. All this activity was carried out in order to enhance the performance of education system by effective use of technology. Implementation of system and use of it increased the overall quality of education, enabled to engage teachers and students in techno-savvy environment and to make sure about the productive outcomes from primary schools<sup>[8]</sup>.

"Educational Portal work has established a pilot Portal for Accessing Information. Work has been carried on in one hundred and twenty pilot schools to test the Portal for Accessing Information. Activities related with international projects Global Gateway have been carried out. Within the context of the Basic Education Program, which is carried on with the support of the World Bank, IT classes will be established in 15 000 schools in rural areas, coordinators of 18 000 IT classes will be educated, 200 000 educational staff will be given in-service training on computer literacy and computer assisted education. And within the context of this program 51 465 computers were provided for 26 276 primary schools and a printer, a scanner, software and an uninterruptible power supply (UPS) were provided for every one of these 26 276 schools. Within the context of Secondary Education Project, which is at the preparatory stage, activities will be carried out to increase the duration of compulsory education to 12 years in the long term, increase the ratio of vocational and technical education in secondary education, develop vocational and technical educational programs and reach EU standards in terms of educational programs"<sup>[9]</sup>.

"The project also aimed to restructure vocational guidance and orientation system at secondary education, using modular teaching methods and educational technologies at secondary education, increase the quality at educational institutions, and determine regional needs in parallel with planned school development model"<sup>[9]</sup>.

### III. METHODOLOGY

#### A. Data Collection

To collect the data an integrated system that collects the student data as he/she enrolls in the education system by joining one of registered institutions has been used. As student progresses all the data relating to attendance, exam results, courses etc. updates it- self in the same repository from time to time. This system can be useful in many regards, since real time data of student is collected and processed. In this study we will be focusing on the data of school pass outs for last three years who took admissions in colleges either in public sector or private sector institution. In total 6 colleges and 4 universities are considered to find the trends relating to new admissions and drop outs.

Figure 2 represents the features (actions) relating to the proposed system to collect and manage all the information from institutions along with all possible stakeholders that can access the system.

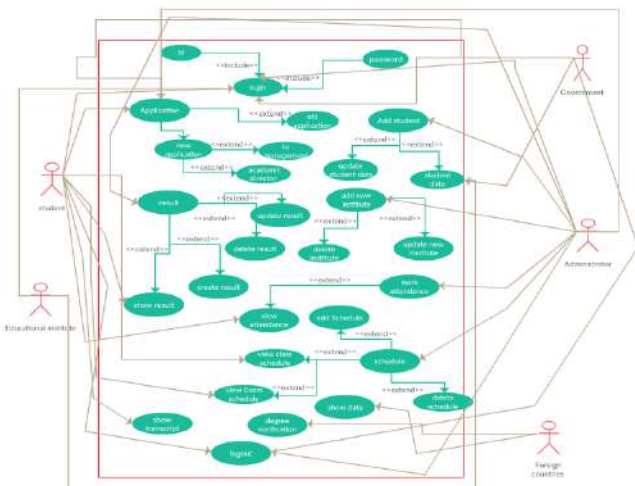


Fig 2

Figure 3 represents the ERD (Entity Relationship Diagram), backend design for the system to collect data from institutions has been represented:

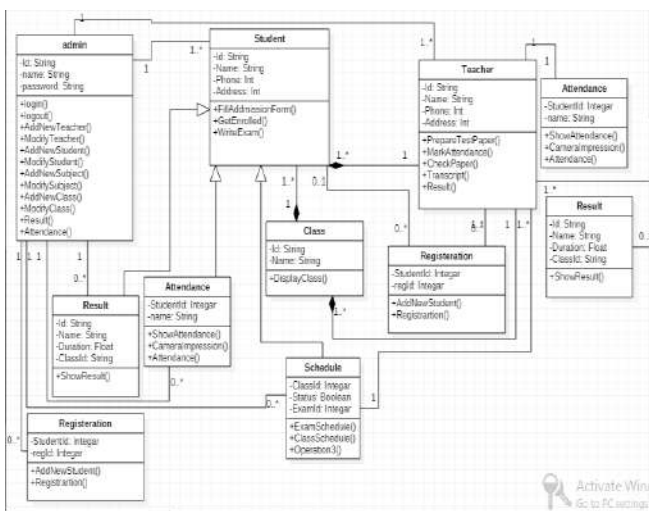


Fig 3

#### B. Machine Learning Tool

Tool for analysis of collected data that is used in this study is WEKA(Waikato Environment for Knowledge Analysis). WEKA is one of the most adapted tools for data analysis these days facilitating researchers to apply of different Machine learning algorithms. It provides the prebuilt functionality to apply most of algorithms using classification techniques and forecasting.

#### C. Data Processing

In pursuance of processing data collected in a data bank that is a centralized repository of all the focused institutions, reports relating to admissions, examinations, attendance etc. are generated in CSV (Comma Separated Values) format. In the next step, CSV file is converted into ARFF (Attribute Relationship File Format) that makes the collected data executable in WEKA. There are normally two types of data, numeric and nominal, in this study we have focused only on the numeric data set of new admissions for last three years in both public and private sector colleges and universities. ARFF contains three segments; first being the relation can be termed as the title of entity (object), other are the attributes (properties) representing the title of columns of CSV file, where the nature of data (either nominal or number) has to be defined. Possible values in each attribute should be marked in case the attribute possess nominal data (for example @attribute Gender {Male, Female}) and marked as real in case the data is numeric for that particular attribute. Third parameter in ARFF is actual data that is collected by the system and is converted as comma separated values.

Although data bank (central repository) contains all the academic data of a student from pre-primary schooling to higher level of education, but we are concerned with the admission and graduation data set of colleges and universities for last three years.

It is important to note that irrelevant or unreliable data makes the knowledge discovery process difficult that is why defining the structure to collect data at back end should be normalized. Data preparation and filtering steps can take a considerable amount of processing time.

Real time data gathering should be smooth in order to facilitate the users to read and write the facts relating to education. Academic and non-academic staff of selected institutions have been trained how to use the data bank so that real values are collected. Student data of admission batch wise have been considered whereas number of drop out students over last 3 years irrespective of their date of enrolment has been selected for processing.

#### D. Data Analysis

Dataset relating to admissions for last three years from 6 colleges and 4 universities has been selected to carry out this research. Three of the 6 colleges are administered by local Governments (stating them as public sector colleges) and remaining 3 are managed by private parties / bodies (termed as private sector colleges). Similarly 2 of the 4 selected Universities are public sector



and other 2 are private sector institutions. Focused dataset has been processed using Gaussian processes - Machine learning algorithms to predict the future trends.

“Gaussian distributions are useful for modelling finite collections of real-valued variables because of their nice analytical properties. Gaussian processes are the extension of multivariate Gaussians to infinite-sized collections of real valued variables. In particular, this extension will allow us to think of Gaussian processes as distributions not just over random vectors but in fact distributions over random functions”.

### IV. RESULTS

After the analysis of data processed using WEKA by applying Gaussian processes, following trends have been generated. Based on the training set of values of last three years these trends are generated.

The application of Gaussian processes on collected data set for admissions and drop outs for public and private sector institutions gives the following results:

Figure 4 depicts the future trends for admissions and drop outs for public sector colleges:

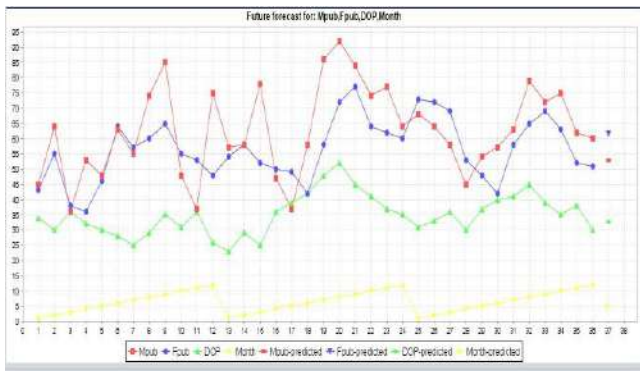


Fig 4

This predicts that the drop out ratio in public sector colleges is on higher side as compared to admissions.

Figure 5 represents the future trends for admissions and drop outs for private sector colleges

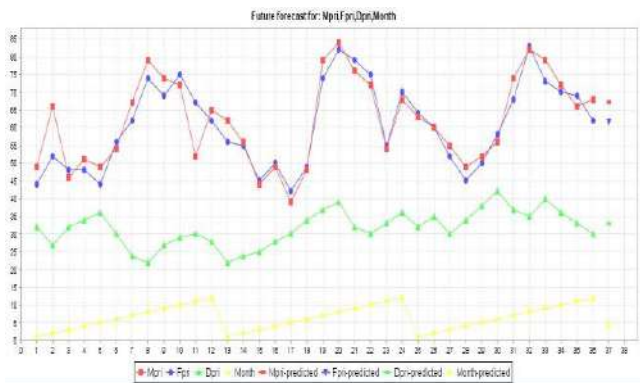


Fig 5

Figure above represents that private sector colleges have a low drop out ratio as compared to admissions.

Figure 6, trends of admissions and dropouts of public sector universities are represented:

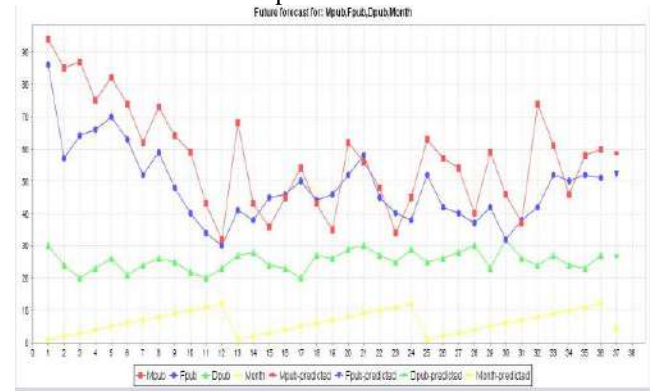


Fig 6

Figure 7 displays the similar trends for private sector universities.

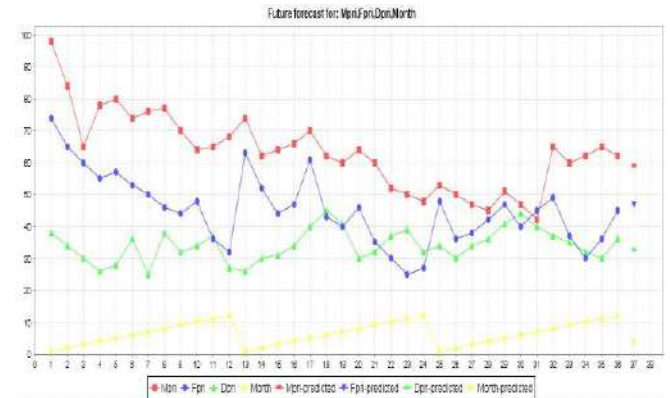


Fig 7

It has been observed that the predicted dropout ratio as compared to new admissions is on higher side in public sector colleges where as it is likely to retain major population in private sector colleges. Vice versa in universities where private sector has more chances of dropouts rather than the public sector universities. Reason may be the fee structure and affordability for higher education.

### V. CONCLUSION

This study helps to propose a system that can enable data collection process from educational institutions at all levels and can help to find the future trends in any of the segment. This paper focusses on the data set relating to new admissions and drop out for students after completing secondary schooling. Hence comparing the performance of public and private sector colleges and universities by analyzing the trends of admissions and number of graduating students, identifying the drop out ratio in each sector. Collected data can further be analyzed in different aspects to find factors effecting performance of students and institutions.

## REFERENCES

- [1]. Citation: Farooq MS (2015) MDGs and Quality Education Situation at Primary Level in Pakistan. *Social Crimonol* 3: 126. doi:10.4172/2375-
- [2]. Citation: Farooq MS (2015) MDGs and Quality Education Situation at Primary Level in Pakistan. *Social Crimonol* 3: 126. doi:10.4172/2375-
- [3]. Citation: Farooq MS (2015) MDGs and Quality Education Situation at Primary Level in Pakistan. *Social Crimonol* 3: 126. doi:10.4172/2375-
- [4]. Farooq MS (2015) MDGs and Quality Education Situation at Primary Level in Pakistan. *Social Crimonol* 3: 126
- [5]. Balanskat, A., Blamire, R., & Kefala, S. (2006). A review of studies of ICT impact on schools in Europe. *European Schoolnet.Bingimlas*,
- [6]. K.A. (2009). Barriers to the Successful Integration of ICT in Teaching and Learning Environments: A Review of the Literature. *Eurasia Journal of Mathematics, Science and Technology Education*, 5(3), 235-245.
- [7]. Maignan, C.J., Ottaviano, G. I. P., & Pinelli, D. (2003). ICT, Clusters and Regional Cohesion: A Summary of Theoretical and Empirical Research. FEEM Working Paper No.58.2003.
- [8]. Flecknoe, M. (2002). How can ICT Help us to Improve Education? *Innovations in Education and Teaching International*, 39(4), 271-279.
- [9]. Fredriksson, U., Gajek, E., & Jedeskog, G. (2009). Ways to use ICT in schools to optimize the impact on teaching and learning. *ActaDidacticaNapocensia*, 2(4), 21-32.
- [10]. Gichoya, D. (2005). Factors Affecting the Successful Implementation of ICT Projects in Government. *Electronic Journal of e-Government*, 3(4), 175-184.
- [11]. Holmes, B., Tangney, B., Fitzgibbon, A., Savage, T., & Mehan, S. (2001). Communal Constructivism: Students constructing learning for as well as with others, in J. Price, D. Willis, N.E. Davis & J. Willis (Eds). *Proceedings of the 12th International Conference of the Society for Information Technology and Teacher Education (SITE 2001)*, 3114-3119, 5-10 March, Orlando.
- [12]. Jamtsho, S., & Bullen, M. (2007). Distance Education in Bhutan: Improving access and quality through ICT use. *Distance Education*, 28(2), 149-161.
- [13]. Kirschner, P., & Davis, N. (2003). "Pedagogic benchmarks for information & communications technology in teacher education". *Technology, Pedagogy and Education*, 12(1), 125-148.
- [14]. Kozma, R. B. (2005). "National Policies That Connect ICT-Based Education Reform to Economic and Social Development". *Human Technology*, 1(2), 117-156.
- [15]. Krishnaveni, R., & Meenakumari, J. (2010). "Usage of ICT for Information Administration in Higher Education Institutions – A study". *International Journal of Environmental Science and Development*, 1(3), 282-286.
- [16]. Osborne, J., & Hennessy, S. (2003). "Literature review in science education and the role of ICT: Promise, problems and future directions". London: Futurelab.
- [17]. Richards, C. (2010). "From old to new learning: global imperatives, exemplary Asian dilemmas and ICT as a key to cultural change in education". *Globalization, Societies and Education*, 2(3), 337-353.
- [18]. Seely, B.E.(2003). "Historical Patterns in the Scholarship of Technology Transfer". *Comparative Technology Transfer & Society*, 1, 7-48.
- [19]. Sein, M.K., & Harindranath, G. (2004), "Conceptualizing the ICT Artifact: Toward Understanding the Role of ICT in National Development". *The Information Society: An International Journal*, 20(1), 15-24.
- [20]. Tearle, P. (2004). "A theoretical and instrumental framework for implementing change in ICT in education". *Cambridge Journal of Education*, 34(3), 331-351.