TIME SERIES ANALYSIS AND FORECASTING IN PAKISTAN EDUCATION MANAGEMENT INFORMATION SYSTEM USING DATA MINING TECHNIQUES

By

Muhammad Bilal Kakli
2010-NUST-MS PhD-CSE(E)-07
MS-10 (CSE)

Submitted to Department of Computer Engineering
in fulfillment of the requirements for the degree of
Masters of Science
in
Computer Software Engineering

Thesis Supervisor
Dr. Usman Qamar

College of Electrical and Mechanical Engineering
National University of Sciences and Technology
June 2013

ACKNOWLEDGEMENTS

First and foremost, I would like to thanks ALLAH ALMIGHTY Whose blessings give me the confidence to complete this task.

I wish to express my profound love and affection to my family specially my mother who brought up our family through very challenging circumstances after the demise of my father. Her prayers gave me the strength to achieve this level.

Besides this, I would like to thank my advisor Dr. Usman Qamar for his encouragement and guidance throughout the thesis period. I appreciate all his contributions of time, ideas, discussion, and critics etc. to make my research experience productive.

Sincere thanks are offered to my NEMIS colleagues: Zubair Farooq, Adeel Zia and Imran Javed for their continuous support.

This page is intentionally left blank.

To my beloved family, and NEMIS colleagues

ABSTRACT

The educational planners need comprehensive, reliable, up-to-date, and timely data for decision making purposes, related to planning, monitoring, evaluation, and review of education system. Such education related data collection and representation occurs at all levels, from grass root level to higher education level.

To facilitate policy makers, educational planners and managers at all administrative levels in policy formulation and management of education system, a computerized educational management information system (EMIS) was established in Pakistan. EMIS functioning at federal level is called National Educational Management Information System (NEMIS). NEMIS has the responsibility of consolidation of national education data and generates useful information to cater needs of policy planners and decision makers. The NEMIS strive to improve the existing decision making system by adopting some of the modern trends in information analysis and using suitable technologies available for this purpose. However, currently NEMIS lacks the capacity to utilize the technologies which are available for analysis and decision making.

This study is aimed to investigate the existing shortcomings faced by Policy and Decision Makers of EMIS and to prepare an automated tool for policy makers for enhancing the efficiency and effectiveness of EMIS. In this study, the data mining concepts and techniques i.e. Time Series Analysis and forecasting will be applied on the existing EMIS data. This will assist education managers to diagnose the issues and problems in the education system and to evolve remedial measures for improvement of the education system.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

Abbreviation	**Illustration**
ARMA	Auto Regressive Moving Average
AR	Auto Regressive
MA	Moving Average
ARIMA	Auto Regressive Integrated Moving Average
UN	United Nations
AEPAM	Academy of Educational Planning and Management
GoP	Government of Pakistan
EDO	Executive District Officer
EMIS	Education Management Information System
NEMIS	National Education Management Information System
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations Children's Fund (formerly United Nations International Children's Emergency Fund)
FATA	Federally Administered Tribal Areas
GB	Gilgit-Baltistan
ICT	Islamabad Capital Territory
KP	Khyber Pakhtunkhwa
AJK	Azad Jammu and Kashmir
NWFP	North West Frontier Province
FANA	Federally Administered Northern Areas
ASC	Annual School Census
EFA	Education for All
MDG	Millennium Development Goal
UNDP	United Nations Development Programme
UIS	UNESCO Institute of Statistics
GMR	Global Monitoring Report
GES	Global Education Digest
HIV/AIDS	Human Immunodeficiency Virus infection / Acquired Immunodeficiency Syndrome
GIR	Gross Intake Rate
NIR	Net Intake Rate

Abbreviation	**Illustration**
GER	Gross Enrolment Ratio
NER	Net Enrolment Ratio
OOSC	Out of School Children
UPE	Universal Primary Education
TR	Transition Rate
PR	Promotion Rate / Percentage of Repeaters
RR	Repetition Rate
DR	Dropout Rate
SR	Survival Rate to Grade 5
CE	Co-efficient of Efficiency
YIG	Years Input per Graduate
PTR	Pupil Teacher Ratio
PCR	Pupil Classroom Ratio
EPSSim	Educational Policy and Strategy Simulation
EdStats	Education Statistics
ANPRO	Analysis and Projection Model
EPDC	Education Policy and Data Center
EDSS	EMIS-Decision Support System
GIS	Geographical Information System
NIPS	National Institute of Population Studies
ADO	Assistant District Officer
TechCOM	Technical Committee
SQL	Structured Query Language
GUI	Graphical User Interface
NIPS	National Institute of Population Studies

Chapter 1
INTRODUCTION

1.1 Overview

Education is considered as the basic fundamental right of every human being. The United Nations (1948) in its Universal Declaration of Human Rights has defined the education as the right of every human being [1]. The developments in the world are evident of the importance of education. Education is used as a tool to enhance the productivity of individuals which ultimately contribute to the development of society.

Since its inception, the education system of Pakistan has expanded tremendously, especially during the last couple of decades. As per statistics report issued by the Academy of Educational Planning and Management, there are 258,349 institutions operating throughout the country with enrolment of more than 39 million [25]. Besides the increase in the educational institutions, the education system of Pakistan still face many challenges including lack of access to education, low learning outcomes, and availability of resources in the rural localities.

1.2 Education System in Pakistan

According to the Constitution of Islamic Republic of Pakistan, 1973 as amended to date, Education comes under the purview of Provinces. Since some of the matters were on the concurrent list therefore the Federal and Provincial Governments collectively undertake the education legislation. Among other tasks, the Constitution of Pakistan delegated the Federal Government the responsibility of Policy, planning and promotion of educational facilities [24].

Each province as well as federating unit has Department of education responsible for management of Education in Province. The education ranges from the pre-primary level to technical and higher education.

After the introduction of Local government system through devolution ordinance in 2001, the District government setup has been established all over the Pakistan. Executive District Officer (Education) has the responsibility of Head of Department of Education in the district. District Government has been entrusted with the planning, management and implementation of plans and policies for the School education in line with the federal government’s policies. School education ranges from Pre-Primary to Higher Secondary School Level. The district/local government is now responsible for opening of new schools, providing physical facilities in the schools, monitoring and supervision, and annual evaluation of teachers [26].

All the management levels need reliable and valid information depending upon its functions. Managerial executives make the strategic decisions regarding education system as a whole. Normally these decisions are for long-term planning. Malik argued that "macro-level decisions include strategic, regulatory, and structural policy domains" [17]. Mid-level management concerns with the transformation of these strategic decisions into more technical framework. In Pakistan, managers at this level make ensure that to what extent the objectives of the plans have been achieved.

At the grass root level or the micro level decisions are related to operational tasks, such as daily activities closer to the school. Since decisions at this level are local and immediate in nature hence they require more detailed information [5]. Education district officials are concerned with resource management and control and responsible for ensuring that resources are distributed equitably and efficiently to the various schools in their districts. The local level mangers function is highly dependent on the availability of such information [28].

Not only the availability of information is essential for informed decision making but also it should of a good quality. Lucey has identified characteristics of good information as relevant for its purpose, sufficiently accurate for its purpose, complete enough for the problem being considered, communicated to the right person, communicated in time for its purpose, contains the right level of detail and easily understandable to the users [16]. McHugh has pointed out that quality data has key three attributes which included accuracy, relevancy and timeliness [18].

In conclusion, the three decision making levels need distinct information that ensures the linkages of these levels. Apparently decisions made at each level are complementary; it is, therefore, imperative that the information system should be responsive to fulfill the information needs at each level. The decisions taken at each level require specific information and the main purpose of an educational information system is to provide relevant information to each group in an accurate and timely fashion, and in a form that is usable at the lowest possible. Hua & Herstein have argued that information based decision making in the management of education system has its goal i.e. increase access, efficiency, effectiveness, equity, and quality of education through effective systems of monitoring and evaluation, budgeting and planning, policy research and analysis [11].

The following figure gives a better understanding of the school education system of Pakistan.

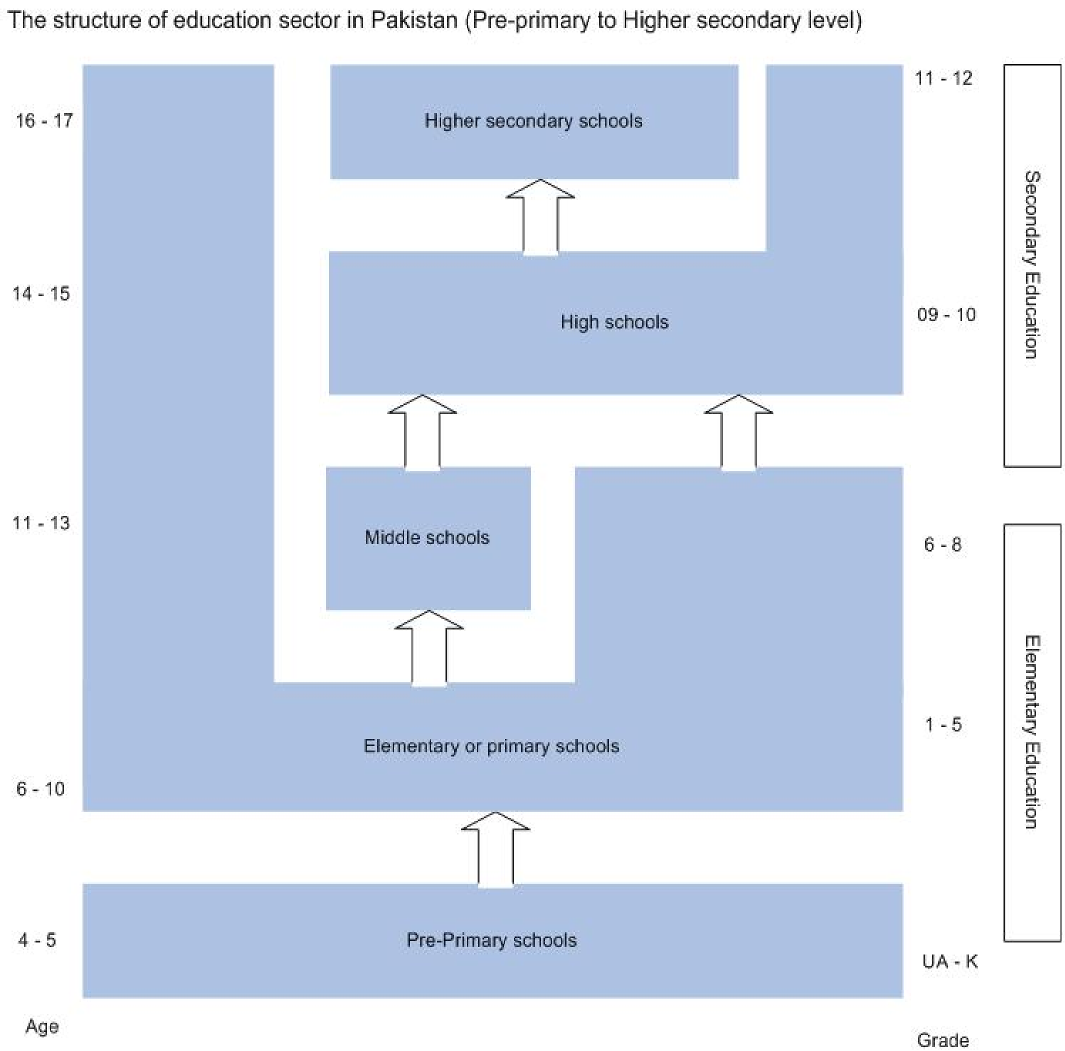


Figure 1.1:*School Education System in Pakistan (Pre-Primary to Higher Secondary)*
(Picture Courtesy Academy of Educational Planning and Management, Islamabad)

1.3 Education Management Information System

In order to accommodate the information needs required for informed policy and decision making at above discussed administrative levels, the establishment and functioning of information system is a must. It is customary that education system like other systems requires information about its inputs, processes and outputs or outcomes. It is obvious that information is needed for diagnosis of current situation of education to identify issues, to set priorities, formulate policies, plans, strategies and to monitor plan implementation [24].

Therefore, information system is a foundation for management of an education system.

According to Laudon & Laudon, an information system is a set of interrelated components that collect (or retrieve), process, store, and distribute information to support decision making and control an organization [14]. Lucey has defined management information system as "a system using formalized procedure to provide management at all levels in all functions with appropriate information, based on data from both internal and external resources, to enable them to make timely and effective decisions for planning, directing and controlling the activities for which they are responsible" [16]. Management information system is a computer-based system that makes information available to the users.

McHugh has expounded that information system is an open system because all its components are working together to process data and produce information. So education is a loosely-coupled or open system [18].

The loosely-coupled nature of education systems makes it hard to organize many things in education including EMIS. One of the challenges for us in setting up a highly systematic EMIS is doing so in such a loosely-coupled organizational environment.

The establishment of information system has significance in all organizations. The information system caters the needs of organization in terms of information. This information is effectively and efficiently used in informed decision making [5]. In the absence of reliable and valid information, the analysis, planning, implementation, monitoring and supervision cannot be done effectively. This results in non-achievement of objectives. The information system required to maintain the information related to Education system is called Education Management information system.

Education management information system (EMIS) has been defined by the researchers in numerous ways. According to Cassidy, an education management information system (EMIS) is a system for the collection, integration, processing, maintenance and dissemination of data and information to support decision making, policy-analysis and formulation, planning, monitoring and management at all levels of an education system [29]. Khan defined EMIS as “a formally structured data collection, retrieval, processing, and dissemination system used to support the main functions of educational managers and decision makers” [30]. According to UNESCO an educational management information system is an organized group of information and documentation services that collects, stores, processes, analyzes and disseminates information for educational planning and management. EMIS is an integrated human and computer based process for providing information needed to support the activities of the Education Department including management, planning and decision-making.

The functioning of EMIS is dependent upon the organizational structure. Lucey argued that MIS must be appropriate to the organization functions and structure [16]. Khan classified EMIS into two categories one was centralized and the other was decentralized. He described that "in a centralized EMIS model all directions are issued and controlled by one central authority whereas a decentralized EMIS operates in a system where educational decision making takes place at the lower level of bureaucracy" [30]. Either centralized or decentralized EMIS, data is maintained in the database. For managing the database, a database management system is developed. Database management system is created for maintaining the data. A database is an integral collection of computer data, organized and stored in a manner that facilitates easy retrieval.

Researchers agree that a comprehensive EMIS should be capable to maintain both quantitative and qualitative data relating to inputs, processes, and outputs of education system. It keeps data on students, teachers, facilities, examination results, and expenditures. Education data from various sources need to be integrated into EMIS database. Cassidy (2005) described that integration refers to compatibility of data from one source with data from other sources; i.e., that data elements of one type from one source can be easily linked with data of other types from other sources [29]. Hua and Herstein argued that data integration is one of the most important EMIS development strategies. They further added that data from multiple sources, multiple years and levels can be linked, integrated, or merged [11]. So EMIS should be capable to integrate information relating to management of various educational activities and to make it available in a comprehensive way to a variety of users.

1.4 EMIS in Pakistan

In order to accommodate the information needs of policy and decision makers, planners and managers in the field of education, The Government of Pakistan established a computerized Education Management Information System (EMIS) in 1990. It is basically a three-tiered system operating at Federal, Provincial and Districts Levels. Following paragraphs presents the brief overview of EMIS in Pakistan.

National Education management Information System (NEMIS) is functioning in the

Academy of Educational Planning and Management (AEPAM) as federal level EMIS. NEMIS plays the coordinating role among four provinces and federal areas i.e. Azad Jammu and Kashmir, FATA, Gilgit-Baltistan, and Islamabad Capital Territory. The NEMIS collate and consolidate the education statistics at national level. NEMIS is entrusted with the responsibility of maintaining national education database. It also helps provincial and district EMISs for enhancing data generation capacity.

The major source of education information at the national level is National Educational Management Information System (NEMIS). NEMIS maintains national education databases. NEMIS collects and collates education statistics received from four provincial EMIS units in Punjab, Sindh, NWFP and Baluchistan and four federally administered areas Islamabad Capital territory (ICT), Federally Administered Tribal Areas (FATA), Federally Administrative Northern Areas (FANA) and Azad Jammu & Kashmir (AJK). The provinces and federally administered areas (except ICT) receive data from districts through district EMIS cells [24].

EMIS establishment in each province/region is called Provincial/Regional EMIS. Provincial/Regional education departments administer the functioning of these EMISs. The Provincial EMIS collects information regarding school education by conducting school census on annual basis. The annual school census data provides the latest education statistics of the province. Provincial EMISs consolidate the education statistics at provincial level. The consolidated database is then transferred to the NEMIS for consolidation at national level. Since the launching of EMIS in early eighties by the NWFP, the other three provinces established their EMIS units in early 1990 mostly sponsored by donors. Federally Administered Tribal Areas (FATA) had established its EMIS unit in 2002. Federally Administrative Northern Areas (FANA) and Azad Jammu & Kashmir (AJK) had established their EMIS units in 1998 and engaged in collection and compilation of education statistics. To bring all these EMIS units at one platform for compatibility of structure and mechanism, NEMIS was entrusted the responsibilities to serve as a coordinating body and to provide technical and professional supports to these EMISs. NEMIS had provided technical support for the establishment of these units which were responsible for collection, compilation, analysis and dissemination of education statistics/data of their respective provinces [2].

EMIS functioning at district level is called district EMIS cells. These cells are functioning under the administrative control of district governments. The District EMIS cell plays a vital role in conducting annual school census. They are responsible to collect data from public schools through annual school census and to provide district's data to the provinces. The devolution has placed a greater reliance on the districts for collection and compilation of education data. Districts are now responsible for using information in planning and resource allocation.

In view of devolution introduced in 2001, district EMIS cells were established in all districts. However, due to poor infrastructure, lack of operating budget and technical staff, these cells were not in position to fulfill the information needs of district managers. The provincial EMIS units usually have been collecting education data through their district EMIS cells by conducting annual school census in October/November [2].

As mentioned earlier, the education management information system has been operational in the country since 1990 and it has been collecting education data from public schools in respect of number of schools up to secondary level, enrolment, number of teachers and physical facilities. Ahmad & Qureshi described that "education management information system remained limited to collection and collating of data pertaining to educational institutions, students' enrolment, number of teachers and physical facilities. Its analysis and interpretation also remained totally ignorant, resultantly; its use in decision making process was negligible" [31]. The underutilization of EMIS data in Pakistan was described by LeBlanc & LeBlanc as "EMIS units in Pakistan" both individually and collectively are fraught of inefficiencies that render them ineffective and underutilized in education planning and management [32].

By all earlier accounts, despite considerable investment of time and resources, efforts to build effective EMIS in Pakistan have had only limited success and there was little evidence that EMIS has contributed to improving the quality of education in the country. There were many ideas and opinions as to why this was the case. The purpose of this study was to better understand the reasons why efforts to build effective EMIS have not been more successful and to know what needs to be done to assure that on-going efforts to develop effective EMIS will be more successful than has been the case thus far. Therefore, this study was designed to investigate the existing infrastructure of EMIS in Pakistan, its internal dynamics for collecting and managing information, use of information in planning and decision making and to identify issues and problems associated with infrastructure, internal dynamics and with use of data. The main focus of this study was to conduct systematic inquiry aimed at data quality, its use and development of comprehensive EMIS to fulfill information requirements of planners, education managers and decision makers at different levels [24].

Chapter 2

REVIEW OF THE RELATED LITERATURE

The statements about the future are called Projections. The Projections are mostly based on historic data. However, an expert opinion is required for the projections evaluation. Therefore a projections comes to be forecast when the element of judgment is added to the projection. Thus the forecasts depicts more realistic picture of the future. However, the forecasts should be reviewed frequently in order to incorporate the changes that occur with the passage of time. Before discussing the Time Series Analysis and Forecasting, it is better to have the understanding of work done by various organizations contributing in the field of education. In addition to it, an understanding of Education Indicators and currently used methods for enrolment and teachers projection as well as simulation models is necessary.

2.1 Education for All

The Education for All movement was launched by the UNESCO, UNICEF, UNDP and World Bank in 1990 with the aim to achieve universal primary education and remove the illiteracy [46]. Universal Primary Education means every child of primary school age will go to school. These targets were to be achieved by the next ten years. But in 2000, it was found that many countries are still lacking behind. Therefore in year 2000, the international community again defined six goals of education and set the targets to achieve these goals by the year 2015 [48]. Following are the six internationally agreed goals which have to be attained by the 2015:

Goal 1: Expanding and improving comprehensive early childhood care and education, especially for the most vulnerable and disadvantaged children.

Goal 2: Ensuring that by 2015 all children, particularly girls, children in difficult circumstances and those belonging to ethnic minorities, have access to, and complete, free and compulsory primary education of good quality.

Goal 3: Ensuring that the learning needs of all young people and adults are met through equitable access to appropriate learning and life-skills programs.

Goal 4: Achieving a 50 per cent improvement in levels of adult literacy by 2015, especially for women, and equitable access to basic and continuing education for all adults.

Goal 5: Eliminating gender disparities in primary and secondary education by 2005, and achieving gender equality in education by 2015, with a focus on ensuring girls' full and equal access to and achievement in basic education of good quality.

Goal 6: Improving all aspects of the quality of education and ensuring excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy and essential life skills.

Pakistan is also one of the countries who agreed for achieving these goals by year 2015. Since National Education Management Information system (NEMIS) is the official source of education statistics, therefore, NEMIS calculates the indicators defined in these goals on yearly basis.

It is expected from the participating countries that they would strive hard and make plans, procedures, develop the capacity, and mobilize the resources in order to achieve these goals. Each goal contains different indicators and the target has been set for each indicator which have to be achieved by 2015. The Education for All also prepared and issued the technical guidelines for calculation and limitations of the indicators. The indicators discussed and calculated in present study are part of EFA Goals.

The UNESCO in collaboration with UIS produces the yearly report namely "Education for All Global Monitoring Report". This report presents the detailed analysis of current situation of EFA targets of each member country [50]. Following are some lines from the EFA Global Monitoring Report 2012:

> "With just three years to go until the deadline for the Education for All goals that were set in Dakar, Senegal, it is vitally urgent to ensure that the collective commitments made by 164 countries in 2000 are met. Lessons also need to be drawn to inform the definition of future international education goals and the design of mechanisms to make sure that all partners live up to their promises. Unfortunately, this year's EFA Global Monitoring Report shows that progress towards many of the goals is slowing down, and that most EFA goals are unlikely to be met. Despite the gloomy outlook overall, progress in some of the world's poorest countries shows what can be achieved with the commitment of national governments and aid donors, including greater numbers of children attending pre-school, completing primary school and making the transition to secondary education [50]."

2.2 Millennium Development Goals

The Millennium Development Goals are eight international development goals that were established by the United Nations in year 2000 [44]. All the member states of United Nations agreed to achieve these goals by year 2015. Following are the eight Millennium Development Goals:

1. Eradicating extreme poverty and hunger,
2. Achieving universal primary education,
3. Promoting gender equality and empowering women,
4. Reducing child mortality rates,
5. Improving maternal health,
6. Combating HIV/AIDS, malaria, and other diseases,
7. Ensuring environmental sustainability, and
8. Developing a global partnership for development.

Each goals have clearly defined targets and the date to which these targets have to be achieved. Two out of these eight goals specifically deals with Education with emphasis on Primary Education and Literacy.

Pakistan also committed to achieve these goals and targets by the year 2015. The Planning Commission of Pakistan yearly publishes report on the current status of Millennium Development Goals. Following are some lines from the Pakistan Millennium Development Goals Reports:

> "The fourth in the series, the Pakistan Millennium Development Goals Report 2010 covers the four years since 2006 in which numerous and far reaching developments have taken place which have transformed the social, political, and economic landscape of Pakistan, all having an impact on the outcomes, achievements and targets of Pakistan's Millennium Development Goals. Pakistan has faced serious challenges in the last four years towards meeting many of the MDG targets; stemming largely from a sudden meltdown in the global economy in 2008, along with a sharp rise in oil and food prices earlier that year. Additionally many serious political and economic problems, both external and domestic, have affected developments since 2006. The militancy, the political instability in 2007-2008 and the transition from a military-led regime to a democratically elected government caused severe disruptions in economic and social development. Furthermore, the most recent catastrophic flood, has affected approximately more than 12 million people, ravaged different rural and urban areas and caused immense damage to the infrastructure and agriculture of the country. This will adversely impact the overall economy and the achievement of many of the MDG goals and targets over the next few years. [43]"

2.3 Education Indicators

An Indicator is the quantitative evidence that point towards some situation. Indicators have been defined by researcher in a number of ways. According to the Seers, "The most important use of development indicators is to provide targets for planning". Cohen 1967 stated that “Accurate meaningful indicators are a Prerequisite for valid answers and for planning affective remedies when needed”

Coombs, 1969 argued that one of the necessary steps towards identifying the right question… is to developed batter instruments for assessing the performance of education system and their varied parts…. What sorts it indicators (does) an education system needs to give itself an annual checkup? Without battery of indicators of an appropriate sort educational planners and policy makers have little to go on except hunch and prejudice [53].

Education Indicators means indicators which are related to education system. These indicators plays significant role in evaluation of present state of education system. The education indicators depict the current situation of education system to the policy and decision makers [51]. These indicators provide the baseline for future policy and planning decision.

The following are the characteristics of a good indicator [51]:

(a) An indicator should provide useful information to the policy makers
(b) Its ability to summarize information without distortions
(c) Its precision and comparability
(d) Its reliability and frequency of updating
(e) It allows relating it with other indicators for global analysis
(f) It measures how far or how close one is from the objectives
(g) It helps to identify problematic or unacceptable situation
(h) It meets policy concerns; and
(i) It helps to compare its value to a reference value, to a norm/standard or itself, as computed for different periods

Following are the major indicators which are calculated in this study.

2.3.1 Gross Intake Rate

It is also known as ‘admission rate’ or ‘entry rate’. It is a very important indicator in knowing the coverage of the entry age of the child population in the system. It indicates the access of primary school education to the population.

It is defined as the new entrants (irrespective of age) to Primary Grade 1 expressed as a percentage of Population of official school entrance age. [Technical Guidelines] It is calculated by dividing the number of new entrants in Grade 1 to the population of age-5 being the official school entrance age. Following is the formula of calculating Gross Intake Rate [52]:

$$GIR^t = \frac{N^t}{P_a{}^t} * 100 \qquad \text{Equation … 2.1}$$

Where:

GIR^t Gross Intake Ratio in School year t

N^t Number of new entrants in the first grade of primary education, in school year t

$P_a{}^t$ Population of official primary school entrance-age a, in school year t

It is pertinent to mention here that when data on new entrants is not available then new entrants are calculated by subtracting the repeaters from the total enrolment.

A high number of Gross intake rate shows that high number of school entry age children has access to primary education.

2.3.2 Net Intake Rate

The NIR is used to precisely measure access to primary education by the eligible population of primary school-entrance age. The Net Intake slightly differs from Gross Intake Rate in the sense that NIR only consider new entrants which are of official primary school entrance age whereas GIR considers all the new entrants irrespective of their age [52].

Following is the formula for calculating the Net Intake Rate:

$$NIR^t = \frac{N_a{}^t}{P_a{}^t} * 100 \qquad \text{Equation … 2.2}$$

Where:

NIR^t Net Intake Rate in School year t

$N_a{}^t$ Number of children of official primary school-entrance age a, who enter the first grade of primary education for the first time in school year t

$P_a{}^t$ Population of official primary school entrance-age a, in school year t

2.3.3 Enrolment Ratio

The enrolment ratio of a specific level is obtained by dividing the enrolment of specific level to corresponding age group population. Two types of enrolment ratios are mostly used. First is Gross whereas second is Net enrolment.

The Gross Enrolment Ratio is calculated by dividing the enrolment (irrespective of age) of specific level by corresponding age group population. The Gross and Net Enrolment Ratios tell the capacity of education system to enroll the students of a particular age group. The gross enrolment rate include both over-age and under-age enrolment. Hence it can exceed hundred percent. Therefore it cannot be used in policy and planning matters [51].

For policy and planning purposes, the Net Enrolment Ratio is used. It is more accurate than GER. The Net enrolment ratio is calculated by dividing the age-specific enrolment of a level by corresponding age group population. Hence the NER cannot exceed hundred percent. If the value of NER exceeds hundred percent then it means that there is something wrong with the data. The higher NER means that more students are enrolled in the specific level of education.

Following are the formulas of GER and NER respectively:

$$GER_h^t = \frac{E_h^t}{P_{h,a}^t} * 100 \qquad \text{Equation … 2.3}$$

Where

GER_h^t Gross Enrolment Ratio at level of education h in school year t

E_h^t Enrolment at the level of education h in school year t

$P_{h,a}^t$ Population in age group a, which officially corresponds to the level of education h in school year t

$$NER_h^t = \frac{E_{h,a}^t}{P_{h,a}^t} * 100 \qquad \text{Equation … 2.4}$$

Where

NER_h^t Net Enrolment Ratio at level of education h in school year t

$E_{h,a}^t$ Enrolment of the population of age group A at level of education h in school year t

$P_{h,a}^t$ Population in age group a, which officially corresponds to the level of education h in school year t

2.3.4 Out of School Children

The Enrolment ratios are very significant in calculating Out of school children. Out of school children are those children which are never admitted in a school or which have not access to school. The out of school children of a specific level are calculated by subtracting the age-specific enrolment of that level from the corresponding population of that age-group.

Another method of calculating the out of school children is by using Net Enrolment Ratio. In this method, if net enrolment ratio is 80 percent then it means that 20 percent children are out of school.

Presently, many international organization and agencies are focusing on out of school children. According to the study by UNESCO, 5.1 million children of primary age group are out of school in Pakistan [50]. It is also the goal of Education for All Forum to attain Universal Primary Education by 2015 which means that no children will be out of school.

2.3.5 Transition Rate

Transition Rate is defined as percentage of students admitted to the first grade of higher level of education to the number of pupils enrolled in the final grade of lower level of education in the previous year [52]. This indicator is used to measure the percentage of students which shift from one level education to the next. It is calculated by dividing the number of new entrants in the first grade of the higher level of education by the number of pupils of final grade of previous year multiplies by 100.

Following is the Formula for calculating Transition Rate:

$$TR_{h,h+1}^{t} = \frac{E_{h+1,1}^{t+1} - R_{h+1,1}^{t+1}}{E_{h,n}^{t}} * 100 \qquad \text{Equation … 2.5}$$

Where:

$TR_{h,h+1}^{t}$ Transition rate (from cycle or level of education h to h+1 in school year t)

$E_{h+1,1}^{t+1}$ Number of pupils enrolled in the first grade at level of education h+1 in school year t+1

$R_{h+1,1}^{t+1}$ Number of pupils repeating the first grade at level of education h+1 in school year t+1

$E_{h,n}^{t}$ Number of pupils enrolled in final grade n at level of education h in school year t

A High Transition Rate shows that higher education level has more intake capacity whereas low transition rate specified that the linkage between lower and high level of education is weak which needs to be strengthened.

2.3.6 Flow Indicators

Flow indicators define how the students pass through various stages of education cycle. Presently, various flow indicators are used to assess the education system at various levels. Following are the most commonly used flow indicators:

2.3.6.1 Promotion Rate

It is defined as:

> "Proportion of pupils from a cohort enrolled in a given grade at a given school years who study in the next grade in the following school year [51]."

Promotion Rate tells that how many students are promoted from present grade to next grade. The promotion rate for a specific grade is calculated by the following formula [53]:

$$PR_{i}^{t} = \frac{NE_{i+1}^{t+1}}{E_{i}^{t}} \qquad \text{Equation … 2.6}$$

Where:

PR_i^t Promotion Rate at grade i in school year t

NE_{i+1}^{t+1} New entrants to grade i+1, in school year t+1

E_i^t Number of pupils enrolled in grade I, in school year t

2.3.6.2 Repetition Rate by Grade

After calculating the Promotion Rate, the repetition rate is computed then. Repetition rate is simply the division of number of repeaters in a grade to the enrolment in the previous year but in the same grade. Following formula is used to calculate the Repetition rate

$$RR_i^t = \frac{R_i^{t+1}}{E_i^t} \qquad \text{Equation … 2.7}$$

Where:

RR_i^t Repetition Rate at grade i in school year t

R_i^{t+1} Number of pupils repeating grade I, in school year t+1

E_i^t Number of pupils enrolled in grade I, in school year t

The repetition rate is used to analyze how many students are studying in the same grade in the current school year as compared to the previous school year [51]. The repetition rate effects internal efficiency of the system. The higher number shows the high repetition rate which means that the instructions given to students in school are poor. Ideally the repetition rate should be zero. However this rate is also dependent on the policies of specific areas because some areas or states have the policy of automatic promotion which yield zero repetition rate. Therefore this indicator should be analyzed in conjunction with policies and other indicators of internal efficiency.

2.3.6.3 Drop-out Rate

Another important indicator among the repetition and promotion rate is dropout rate. It is also calculated grade-wise. This indicator tells that how many students are leaving the system grade-wise. Following formula is used to calculate the grade wise dropout rate [52]:

$$DR_i^t = 100 - (PR_i^t + RR_i^t) \qquad \text{Equation … 2.8}$$

Where:

DR_i^t Dropout Rate at grade i in school year t

PR_i^t Promotion Rate at grade i in school year t

RR_i^t Repetition Rate at grade i in school year t

A high dropout rate indicates that most students are leaving the education system. Therefore there is some problem in the internal efficiency of the education system which needs to be addressed.

2.3.6.4 Survival Rate to Grade 5

One of the most important indicators at primary level of education is the survival rate to last grade of primary. The survival rate indicates that how many children who have started grade 1 reached the last grade of primary education.

Different methods have been developed in the past to calculate the survival rate. But the Reconstructed Cohort Model is widely used for calculating the Survival Rate [54]. This method has three assumptions: (i) the drop outs will not enroll again in school, (ii) repetition and drop-out rates will remain constant, and (iii) same rates apply to the new entrants and repeaters.

The Reconstructed Cohort Model required data of enrolment and repeaters by grade for the consecutive two years for computing survival rate.

Following is the formula for calculating the survival rate:

$$SR_{g,i}^{k} = \frac{\sum_{t=1}^{m} P_{g,i}^{t}}{E_{g}^{k}} * 100 \qquad \text{Equation … 2.9}$$

Where:

$P_{g,i}^{t} = E_{g,i+1}^{t+1} - R_{g,i+1}^{t+1}$

i represents grade (1,2,3,…,n)

t represents year (1,2,3,…,n)

g represents pupil-cohort

$SR_{g,i}^{k}$ Survival Rate of pupil-cohort g at grade I for a reference year k

E_{g}^{k} Total number of pupils belonging to a cohort g at a reference year k

$P_{g,i}^{t}$ Promotees' from E_{g}^{k} who would join successive grades I throughout successive years t

R_{i}^{t} Number of pupils repeating grade i in school year t

This indicator is widely used for assessing the internal efficiency of the education system. A high survival rate shows that high numbers of students are retained by the schools whereas low survival rate shows that high numbers of students are dropping out from school. This method is widely used but still it has some limitations in measuring the survival rate as this method does not take into account the re-entrants or migrants from one school to another. Therefore this indicator should be used with other intake and enrolment ratio to have a better sense of internal efficiency and coverage. The details of Reconstructed Cohort Model can be seen at Appendix A.1.

2.3.6.5 Coefficient of Efficiency

This indicator is defined by the UNESCO as following:

> "The ideal (optimal) number of pupil-years required (i.e. in the absence of repetition and dropout) to produce a number of graduates from a given school-cohort for a cycle or level of education expressed as a percentage of the actual number of pupil-years spent to produce the same number of graduates. Input-output ratio, which is the reciprocal of the coefficient of efficiency, is often used as an alternative [52]."

Following is the formula to calculate the co-efficient of efficiency:

$$CE_g = \frac{\sum_{j=n}^{n+k} G_{g,j} * n}{\left\{\sum_{j=n}^{n+k} G_{g,j} * j\right\} + \left\{\sum_{j=1}^{n+k} D_{g,j} * j\right\}} * 100 \qquad \text{Equation … 2.10}$$

Where:

CE_g Coefficient of Efficiency for a pupil-cohort g

$G_{g,n}$ Number of pupils graduating from cohort g in final grade n after n years of study (without repetition)

$G_{g,j}$ Number of pupils graduating from cohort g in final grade n after j years of study

$D_{g,j}$ Number of pupils (of the cohort g) dropping out after j years of study

k Number of repetitions allowed

n Normal duration of study for a cycle or level of education

g Pupil-cohort

j Number of years of study

The value of this indicator shows that efficiency of system to produce graduates. A high value of Co-efficient of efficiency means the system is efficient in producing graduates where as a low number means the students are not completing the level of education due to high number of drop-outs or repetition. This indicator is calculated by the Reconstructed Cohort Model. The care should be observed in using this model as this model is very sensitive to the values of each grade.

2.3.6.6 Years-Input per Graduate

This indicator is used to assess the number of years spend by students who graduate from a specific level of education [51].

Following is the formula to calculate the years input per graduate:

$$\mathrm{YIG_g} = \frac{\left\{\sum_{j=n}^{n+k} G_{g,j} * j\right\} + \left\{\sum_{j=1}^{n+k} D_{g,j} * j\right\}}{\sum_{j=n}^{n+k} G_{g,j}} \qquad \text{Equation … 2.11}$$

Where:

YIG_g Years input per graduate (for graduates belonging to cohort g)

$G_{g,j}$ Graduates from cohort g after j years of study

$D_{g,j}$ Dropouts from cohort g after j years of study

k Number of repetitions allowed

n Normal duration of study for a cycle or level of education

g pupil-cohort

j numbers of years of study

This indicator is also calculated with the help of Reconstructed Cohort Model.

2.3.6.7 Percentage of Repeaters

Repeater in education system is a student who is enrolled in the same grade of education as in previous year. Percentage of such students to the total enrolment of that grade is called the percentage of repeaters [52]. Following is the formula to calculate the percentage of repeaters:

$$PR_i^t = \frac{R_i^t}{E_i^t} * 100 \qquad \text{Equation … 2.12}$$

Where:

PR_i^t Percentage of repeaters in grade i, in school year t

R_i^t Number of pupils repeating grade i, in school year t

E_i^t Number of pupils enrolled in grade i, in school year t

A high value of percentage of repeaters means that there is some problem with the internal efficiency of the education system. Ideally it should be as low as possible.

2.3.7 Pupil-Teacher Ratio (PTR)

In the simples form, the pupil-teacher ratio tells that for how many students, a teacher is available. It is used to measure the number of teachers available to the size of population. It is calculated by dividing the number of students of a specified level to the available teachers at the same level. The higher number shows that more students are accommodated with a single teacher. A low PTR means that class size is smaller and less number of students are taught by a teacher. Hence the teacher can pay more attention to each student within class which results in better performance [53].

Since this indicator is quantitative in nature, therefore it does not consider the teacher qualification, training and experience as well as teaching aids available in the classroom.

Following is the formula for calculating PTR:

$$\mathrm{PTR}_h^t = \frac{\mathrm{E}_h^t}{\mathrm{T}_h^t} \qquad \text{Equation … 2.13}$$

Where:

PTR_h^t Pupil-teacher ratio at level of education h in school year t

E_h^t Total number of pupils or students at level of education h in school year t

T_h^t Total number of teachers at level of education h in school year t

2.3.8 Pupil Classroom Ratio (PCR)

It is defined as the total number of students to which a classroom is available in the school. It is calculated by dividing the number of students of a particular level to the available classroom to that enrolment. A high number of PCR shows that large number of students accommodated in a classroom. Normally the ideal value of PCR is set by the educational authorities.

2.3.9 Indicators of Infrastructure

The availability of infrastructure in the school is very significant as this helps in provision of better learning environment. Many basic and physical facilities can be provided to school depending upon the school level. However, the four physical facilities namely boundary wall, drinking water, electricity, and toilet are considered to be the most important. The following four indicators provide the percentage of schools having these facilities:

1. Percentage of schools having boundary walls
2. Percentage of schools having drinking water facilities
3. Percentage of schools having electricity connection
4. Percentage of schools having toilet facility

These indicators are calculated by dividing the number of schools having particular physical facility to the total number of schools multiplied by 100. The higher percentage shows that large number of schools has the physical facility. These indicators can help the finance authorities in allocating the funds to those schools which lack these facilities.

2.4 Disaggregation of Education Indicators

Disaggregation of an indicator means to break down the indicator to further sub-categories. The disaggregation may point out some variances and gaps that may not be revealed by the overall general figures. Different levels of disaggregation allow us to analyze an indicator with a different perspective and hence the vulnerable areas can be identified. Consider an example of 1500 enrolment (600 male and 900 male) for which 60 teachers are available (40 male and 20 female). If we calculate the overall PTR then it comes to 25 which reflect a very good value. But if this indicator is disaggregated by Gender, then the PTR for become 15 and PTR for female become 45. Hence a more clear picture is obtained which indicates that the female enrolment require more teachers.

The disaggregation of indicators is very important in calculating indicators as it reveals the weak areas which need ample consideration and reforms. The priority areas can also be worked out by disaggregating data at more detailed levels. It is therefore significant to calculate the indicators are various disaggregated levels which results in clearer picture of the education system. Another advantage of disaggregation is that the decision makers can foresee the implication of policies on the vulnerable groups. Hence a better and realistic policy can be defined.

Keeping in view the importance of disaggregation of education indicators, all the indicators discussed in the previous pages are aggregated at the following levels:

- **Gender**
  - Male / Female / Both Genders
- **Administrative Units**
  - Province / District
- **Location**
  - Urban / Rural / Both Locations
- **Level**
  - Pre-Primary / Primary / Middle / High / Higher Sec.
- **Sector**
  - Public / Private / Both Sectors

2.5 Enrolment Projection Techniques

The most important factor of an education system is the student. The number of students in a school/college/university is termed as enrolment. Various methods have been developed for the enrolment projections. Some methods required past five to ten years data whereas some required only previous year data. The projection method is selected based on the available data. These are discussed as follows:

2.5.1 Target Setting Method

This method is used when only the single observation data is available. In terms of enrolment, it is used only when the enrolment data for a single year is available. For application of this method, a target year and the target enrolment is defined based on some assumption. Then the projection is done based on these base and target year.

2.5.2 Simple Growth Rate Method

This is the simplest method for projecting the enrolment. This method requires the enrolment figures for two points i.e. Base year and last year. This method assumes that the growth rate is constant between base year and last year [41]. The growth rate is calculated by the following formula:

$$R = \frac{1}{n}\left(\frac{En - Eo}{Eo}\right) * 100 \qquad \text{Equation … 2.14}$$

Where:

R = annual rate of growth

En= Enrolment in the last year

Eo = Enrolment in the base year

n = number of intermediary years.

On the basis of this growth rate, the projected enrolment for the next years is calculated. This method can be used to project the overall enrolment figures for a level by multiply the last year figures with the calculated growth rate. But this method is normally not used due to the fact that it does not take into account all the values between base year and last year.

2.5.3 Compound Growth Rate Method

In order to overcome the shortcoming of the simple growth rate method, the compound growth rate method was developed. It contains a slight improvement in the simple growth rate method. In this method, the growth rate is calculated by the following formula:

$$R = \left(\left(\frac{En}{Eo}\right)^{\frac{1}{n}} - 1\right) \text{x } 100 \qquad \text{Equation … 2.15}$$

Where

R = annual rate of growth

En = Enrolment in the last year

Eo = Enrolment in the base year

n = number of intermediary years.

Once the growth rate is calculated, the following formula is used to calculate the projection for the next years:

$$En = Eo\left(\frac{1+R}{100}\right)^{n} x\ 100 \quad \text{Equation … 2.16}$$

where

R = annual rate of growth

En = Enrolment in the last year

Eo = Enrolment in the base year

n = number of years.

The major limitation in the above discussed methods is that these do not take into account the complete data. These methods project the figures based on first and last point.

2.5.4 Enrolment Ratio Method

Another shortcoming of the growth rate methods is that these do not take into account the population factor. Therefore the Enrolment Ratio Method was developed and the population was also considered in projecting the enrolment [42].

Following are the steps for calculating a projection using enrolment ratio method:

1) Calculate the population for the required level. For example, if the projection is made for primary level, the population for ages 5-9 is calculated.
2) Calculate the enrolment ratio for each year by dividing the enrolment with population.
3) Project the enrolment ratios for future years by any suitable method.
4) Calculate the enrolment by taking the percentage of enrolment ratio (step 3) to the calculated population (step 1).

The population data varies from time to time depending on Government's health policies, the birth rate and various other parameters. In the absence of accurate population data, the chances of errors occurs as the population projection techniques have their own limitations too. Therefore this method is mostly used when the accurate population data is available.

2.5.5 Method of Least Squares

Method of least squares is another method for projecting the enrolment. This method uses the complete time-series data which is available. This method is also called 'line of best fit'. The method fit the straight line to the time series data [41]. This straight line is extended to make the future projections. In the case when the straight line does not fit, then the method of curve fitting is used. The past time series data is fitted along a curve. The best equation can be identified based on some statistical calculation e.g. Mean Square error or Mean absolute error etc.

The drawback of this graphic method is that different experts will get different figures because this method considers the straight line which touches maximum plotted points.

The above discussed methods i.e. Target Setting, Growth Rate or Method of Least Squares, do not consider the internal dynamics of enrolment. Internal dynamics means enrolment from one grade to another. In order to calculate the grade-to-grade enrolment, the following methods are available:

2.5.6 Grade-Ratio Method

The grade-ratio method uses the following data for calculation of projections:

- Two consecutive years' enrolment and repeaters
- Population (School entrance age) for two consecutive years
- Targets on various rates such as admission, drop-out, promotion etc

When the above mentioned data is available then the projections using the Grade-Ratio methods are made on the basis of following formula:

$$\left(\frac{(E_{t+1} * G_{t+1})}{E_t * G_t}\right) * 100 \qquad \text{Equation … 2.17}$$

Where

E is enrolment

G is Grade

t is year

Before using this method, the care should be observed in selecting the data. If the data is incomplete then the projections will not be reliable [41]. Similarly, the targets for various rates should also be realistic.

In order to obtain the enrolment for the specific level, the sum of enrolment of individual grades is used. For example, the primary enrolment figure will be calculated by adding the enrolment from grade 1 to 5.

2.5.7 Grade-Transition Method

This method is normally used with the repeaters figures are not available. The calculation procedure of Grade-Transition method slightly differs from the Grade ratio method [41]. In Grade Transition Method, the Grade-wise enrolment is calculated by assigning the promotion, repetition and drop-out rates in different grade and years. Previously, only grade-ratios were used in Grade-Ratio method. Therefore this method is more powerful then grade ratio method in the sense that it allows the decision makers to foresee the change in future by changing the factors i.e. repetition, promotion and drop-outs. Hence various assumptions will provide different outcomes which can be used in policy making.

2.6 Projections of Teachers

Teachers are another important factor of an education system. Same as enrolment, the number of teachers are also projected. Following are the methods which are used to calculate the number of teachers:

2.6.1 Pupil-Teacher Ratio (PTR) Method

It is the simplest method used to project the teachers. In this method, the PTR is calculated first by the following formula:

$$R_t = \frac{E_t}{T_t} \qquad \text{Equation … 2.18}$$

Where

T_t Represents Number of teachers at a particular time (t) and for particular level or school

E_t Represents Enrolment at particular time (t) for particular level or school

R_t Represents Pupil-Teacher Ratio at a particular time (t) for particular level or school

When the PTR is calculated, then the projection is made by dividing the enrolment to the PTR. This method assumes that the enrolment means total enrolment of a particular level [41]. This assumption is not true in the rural areas of Pakistan due to multi-grade teaching.

2.6.2 Method Based on Number of Pupil per Class and Hours Taught by Teacher

This method is better than the Pupil-Teacher Ratio Method for calculating the teacher-requirements. This method considers the following variables in making projections:

1) Class size
2) Number of Hours per week student received instruction
3) Number of hours taught per week by a teacher

The following data is required for using this method:

- Stage Wise Enrolment
- Number of hours per week for a student
- Number of students taught at the same time by one teacher
- Number of hours taught by a teacher per week

Following formula is used to make the projections:

$$T = \frac{E * H_s}{R * H_t} \qquad \text{Equation … 2.19}$$

Where

T = Required number of teachers;

E = Projected enrolment;

R = Number of students per teacher or per instructional group or size of class;

H_s = Average number of weekly hours per student,

H_t = Average number of weekly hours per full-time teacher.

The above mentioned formula is very important for decision making because each factor can be changed and the result can be analyzed. Hence the decision maker can foresee the implications of changing one or more factors.

2.7 Projections and Simulation Models in EMIS

2.7.1 EPSSim

EPSSim stands for Education Policy and Strategy Simulation. This model was developed by UNESCO in 2001. The basic aim to develop this model was to enable the member countries to make projections and develop policy to achieve EFA Goals [33]. This tool can be used to develop different scenarios related to education system. These scenarios are then simulated based on policy variables and targets. The data requirement for this model is huge as it needs data about population, enrolment, teaching and non-teaching staff, school infrastructure and classrooms, teaching aids, classroom material etc [34]. Therefore this model can only be used when all this data is available. One major drawback of this model is that it does not take into account the past time series for projections and simulation rather it uses the target setting method, which is explained in the following pages.

In the EPSSim model, the baseline data is entered first. This baseline data can be obtained from database or school records or other statistical sources. On the basis of this data, the targets are defined. Then this model simulates the baseline data to the target year and objectives [33]. Following is the basic diagram of EPSSim Model:

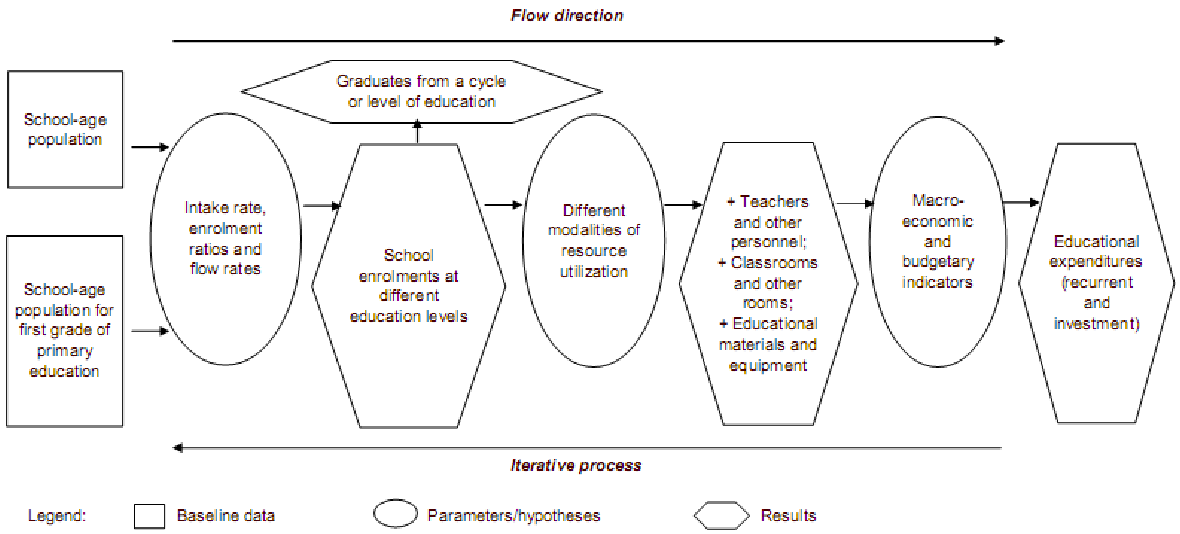


Figure 2.1:*Basic diagram of the simulation flow of EPSSim*

In order to maintain ease of use, this model has been developed in Microsoft Excel program. The macros are defined to perform the simulations. The exercises are also given in the program which helps the users to familiarize with the EPSSim terminology. Following is the screen shot of EPSSim model:

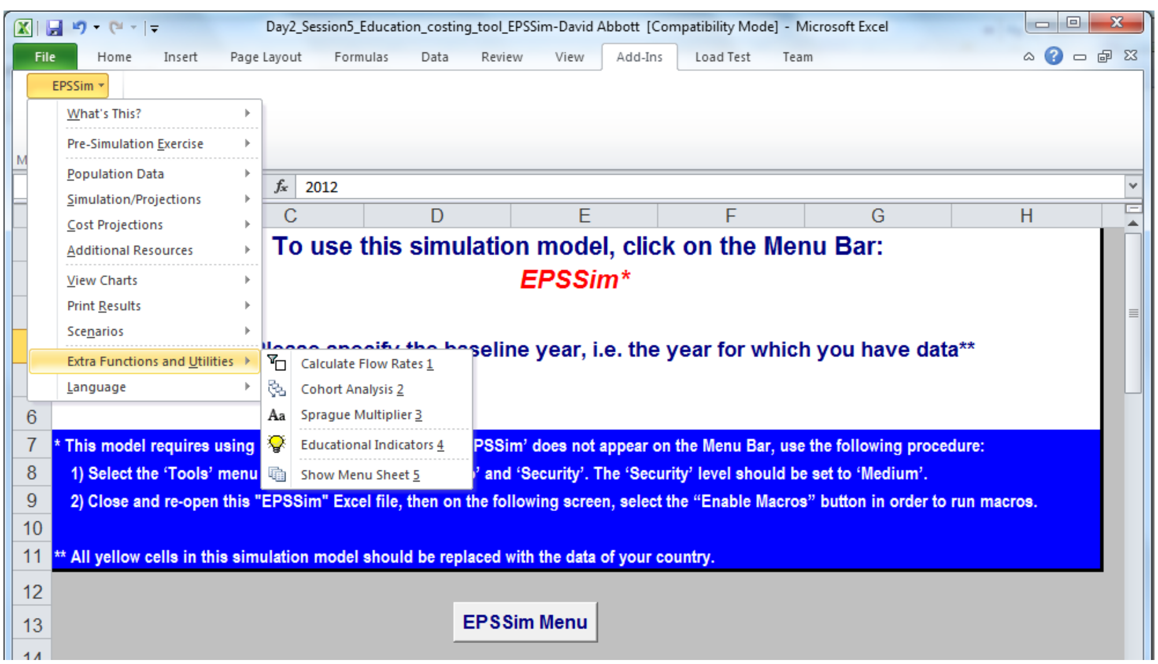


Figure 2.2:*Screenshot showing the startup screen of EPSSim*

2.7.2 EdStats

It stands for Education Statistics. This model was developed by World Bank. This model contains various modules each of which provide the data and projections for the specific portion of education system. This program provides the education projections from year 2005 to 2025. The data was gathered through household surveys, Population and labor projections [35]. The education data is obtained from UNESCO Institute of Statistics. The limitation of this tool is that it makes the projections on the basis of data collected by World Bank's own sources or from UIS. For example, the population projections made by World Bank are used in this program [36]. Therefore the projections made by particular country cannot be used. Special requests can be made to the World Bank for projecting the Non-UIS data such as national figures of particular country but it is a time taking process. Following is the chart produced by EdStats regarding primary enrolment completion rates of Pakistan:

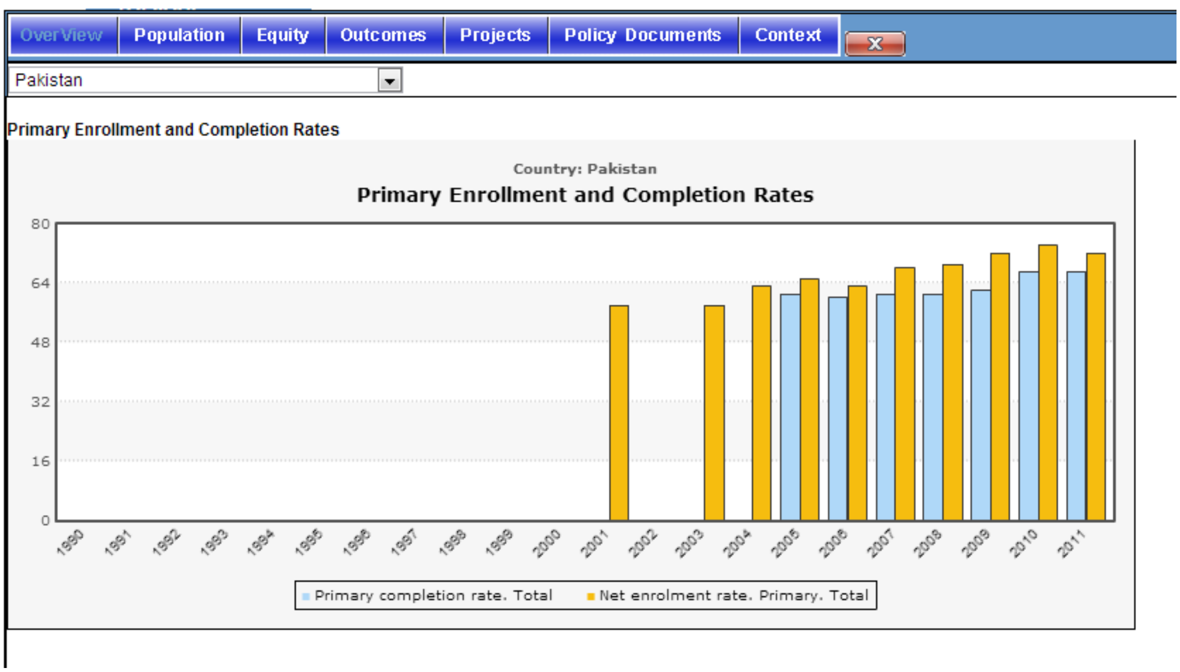


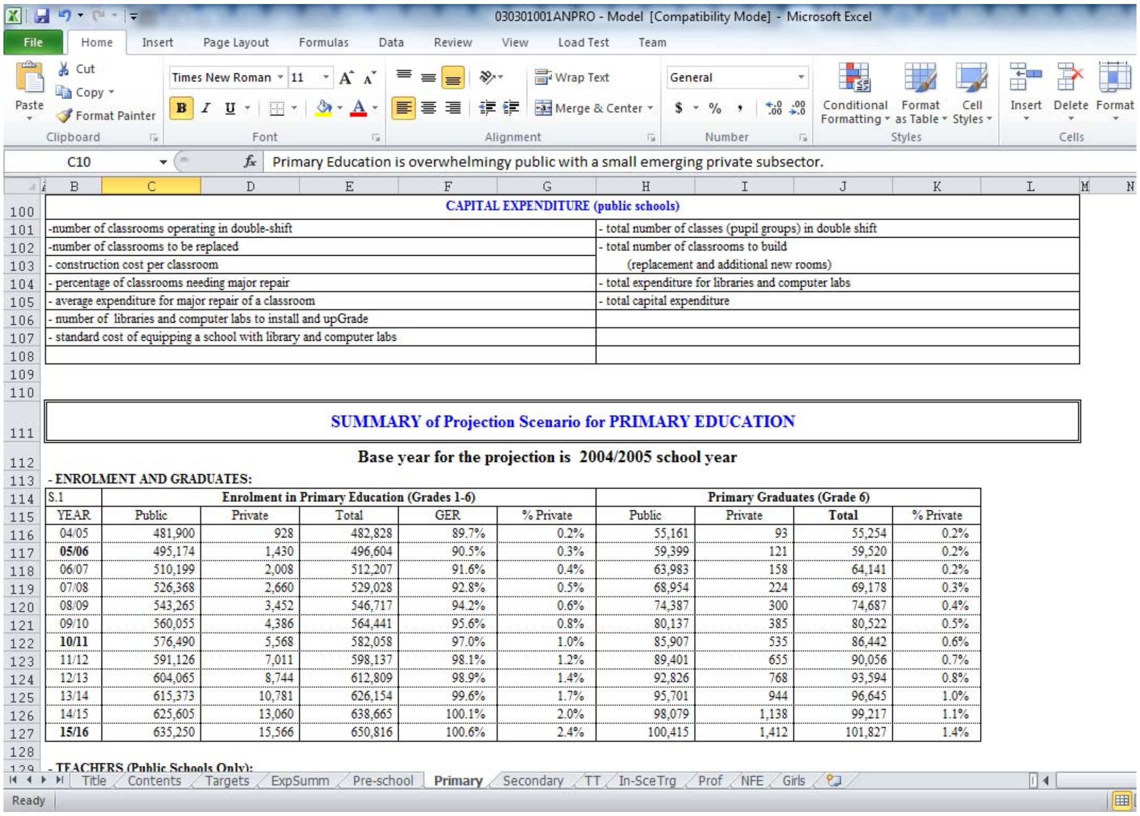
Figure 2.3:*Chart created by EdStats tool*

Another shortcoming of this method is that it does not take into account the changes that occur as a result of policy change.

2.7.3 ANPRO

It is short form of Analysis and Projection Model. It was developed by UNESCO [39]. This model is similar to EPSSim in the sense that it also uses the target setting method. Targets are defined for the particular years and then projections are made on how to achieve the targets

[39]. This Model is slightly limited in nature as compared to EPSSim as it does not deal with the population. But still the data required to use this model is very large.



S.1	Enrolment in Primary Education (Grades 1-6)					Primary Graduates (Grade 6)			
YEAR	Public	Private	Total	GER	% Private	Public	Private	Total	% Private
04/05	481,900	928	482,828	89.7%	0.2%	55,161	93	55,254	0.2%
05/06	495,174	1,430	496,604	90.5%	0.3%	59,399	121	59,520	0.2%
06/07	510,199	2,008	512,207	91.6%	0.4%	63,983	158	64,141	0.2%
07/08	526,368	2,660	529,028	92.8%	0.5%	68,954	224	69,178	0.3%
08/09	543,265	3,452	546,717	94.2%	0.6%	74,387	300	74,687	0.4%
09/10	560,055	4,386	564,441	95.6%	0.8%	80,137	385	80,522	0.5%
10/11	576,490	5,568	582,058	97.0%	1.0%	85,907	535	86,442	0.6%
11/12	591,126	7,011	598,137	98.1%	1.2%	89,401	655	90,056	0.7%
12/13	604,065	8,744	612,809	98.9%	1.4%	92,826	768	93,594	0.8%
13/14	615,373	10,781	626,154	99.6%	1.7%	95,701	944	96,645	1.0%
14/15	625,605	13,060	638,665	100.1%	2.0%	98,079	1,138	99,217	1.1%
15/16	635,250	15,566	650,816	100.6%	2.4%	100,415	1,412	101,827	1.4%

Figure 2.4: *Working Environment of ANPRO model*

One important assumption that this model uses is that the primary education is provided mostly by public sector schools with a very low share of private sector [38]. This assumption is not true in the case of Pakistan as the private share in education system of Pakistan is nearly 40 percent. Therefore this model cannot be applied to the Pakistan’s education system.

2.7.4 EPDC Projections

EPDC is acronym of Education Policy and Data Center. It is part of World Bank. This center has projected the indicators from 2005 to 2050 and these figures are available online [40]. This model also has the same limitation that it uses the population projections of UN/World Bank. Hence national statistics cannot be projected with this method. Another shortcoming of these projections is that these are available only for selected indicators. All indicators are not covered in these projections. Following is the screenshot of EPDC Projections query system:

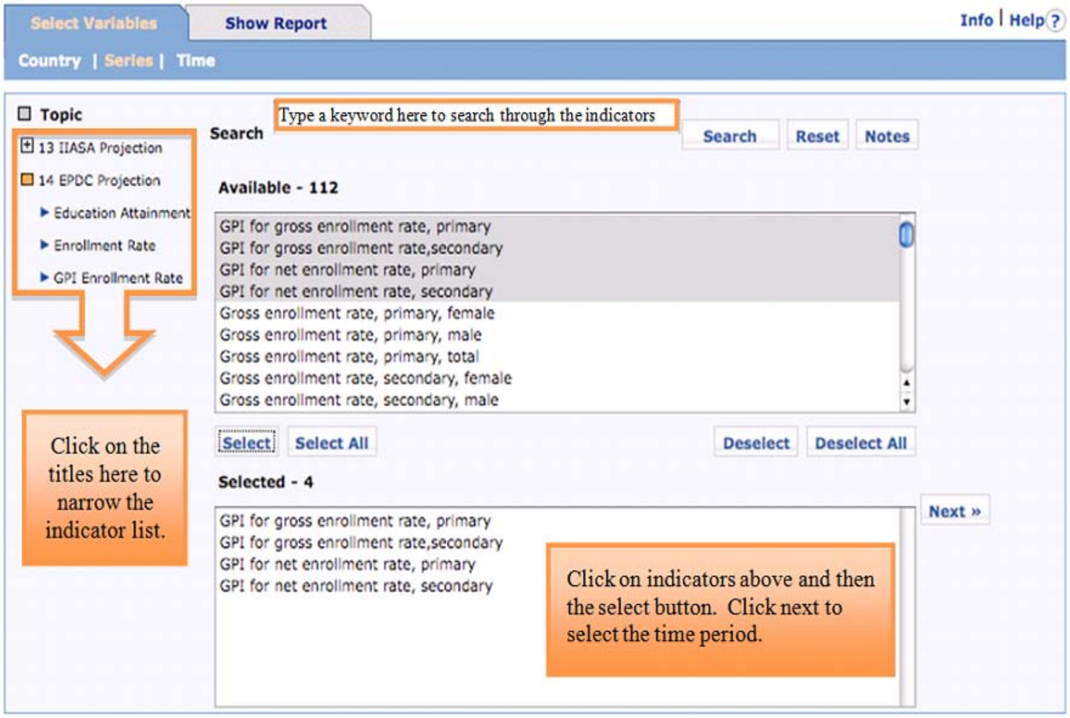


Figure 2.5: *User Interface of EPDC Query System*

2.8 Time Series Analysis

Time Series is arrangement of values that are recorded at consistent time intervals. A Time Series database contains the time series data. With the passage of time, every organization is getting more and more data. Since this data has a time stamp associated with it, therefore this data become time series after some time. This time series become more important when it provide the baseline for the various strategic and policy decisions. The following figure depicts a time series:

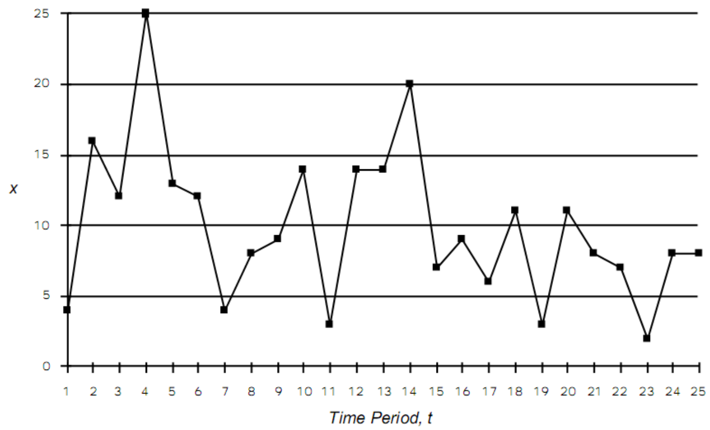


Figure 2.6: *Time Series Graph*

Time Series models are used to analyze the various underlying information about the time series. These models can be applied to a variety of application such as finance, census, education, stock market analysis etc. These time series analysis are further used in forecasting the future values. These forecasted values help the managers and organizations in making their business decisions.

A Time Series can contain various characteristics which include trend, cycle, season, or irregular. If the value of time series increases or decreases with the passage of time then the time series is said to have trend. This time series is also called non-stationary time series and the analysis of such time series is sometimes called trend analysis. Following figure shows the example of trend:

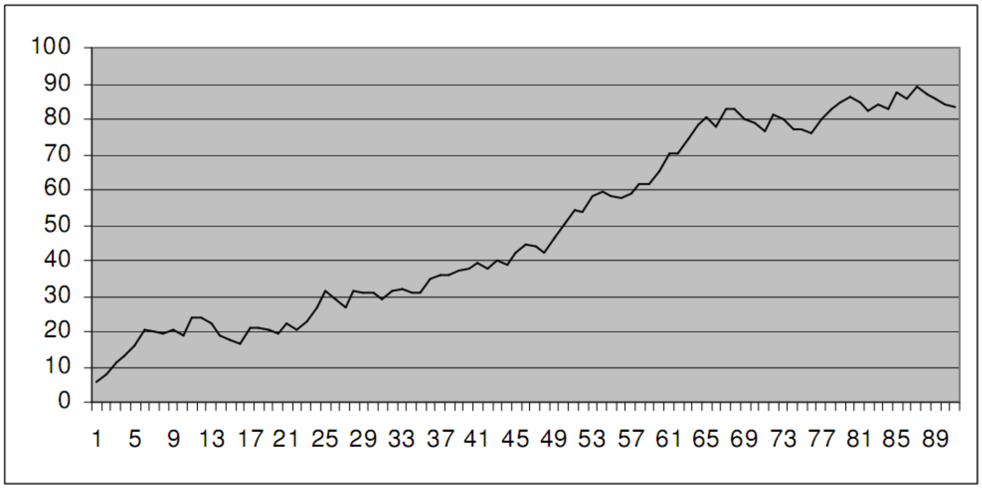


Figure 2.7:*Trend Graph*

A time series is said to have cyclic variation if the values are repeated in reasonably systematic way over time. The cyclic time series are normally represented by wave-shaped graph which represents the increase and decrease. The following figure shows the cyclic time series:

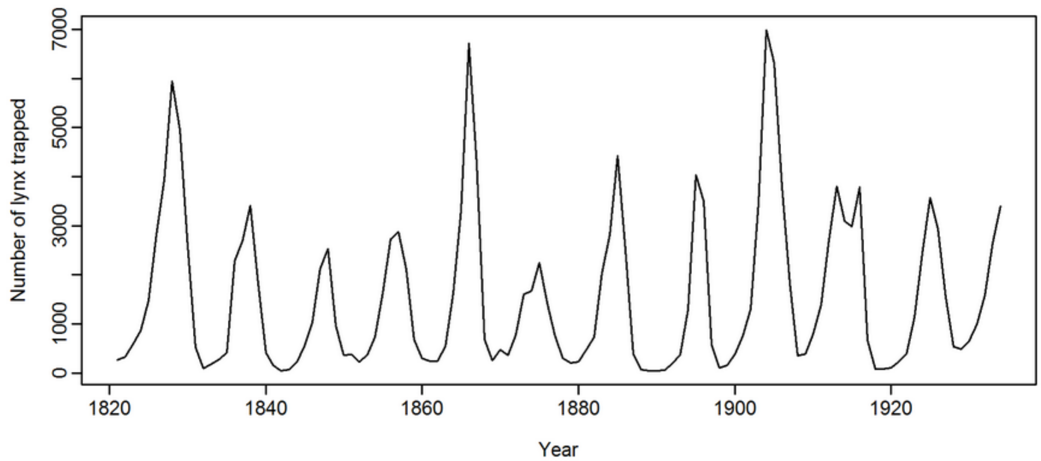


Figure 2.8:*Time series with cyclic characteristics*

Seasonal variations are those variations which occur periodically in the year. These fluctuations go on repeating year by year. The following figure represents a time series having seasonality variation:

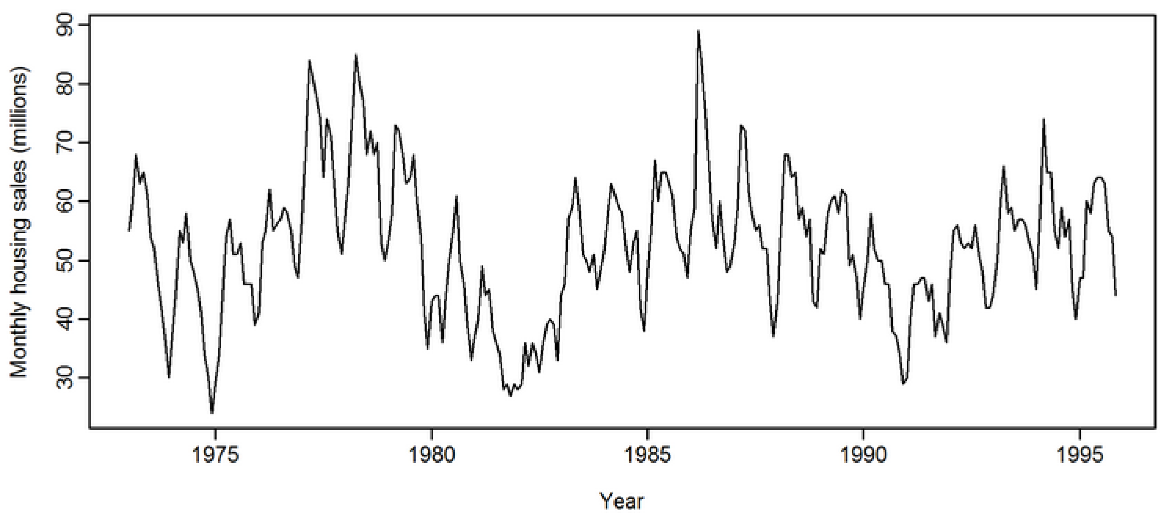


Figure 2.9:*Time Series having seasonality variations*

2.9 Time Series Analysis and Forecasting in Data Mining

Various time series analysis and forecasting methods have been developed in past. In the following sections, the brief overview of each method is presented.

Following are the major Time Series Analysis and Forecasting methods in Data Mining:

2.9.1 Moving average

This method as its name implies takes the arithmetic mean of the series over the complete time series. This method wholly depends on the length of period for calculating arithmetic mean. This method is normally used to smooth the time series. Here is the example of this method:

First Average = MA (3) = $\frac{X1+X2+X3}{3}$ Equation … 2.20

Second Average = MA (3) = $\frac{X2+X3+X4}{3}$ Equation … 2.21

Where MA means Moving Average, X is the time series value and 3 is the length of period for calculating average i.e. 3 years moving average.

2.9.2 Weighted Moving Average / Exponential Smoothing

This method is also used for smoothing the time series data having high variations. This method is based on weighted moving average. In this method, the observations are given weights according to their occurrence. And the weights decline exponentially. The latest observation is given the highest weight. This method is primarily used for short-term forecasting.

The following example shows the working of this method.

An exponentially weighted moving average with a smoothing constanta, corresponds roughly to a simple moving average of length (i.e., period) n, whereaand n are related by:

a= 2/(n+1) OR n = (2 -a)/a Equation … 2.22

Thus, for example, an exponentially weighted moving average with a smoothing constant equal to 0.1 would correspond roughly to a 19 day moving average. And a 40-day simple moving average would correspond roughly to an exponentially weighted moving average with a smoothing constant equal to 0.04878 [55].

2.9.3 Extrapolation

Extrapolation is defined in the Merriam-Webster dictionary as:

> “to predict by projecting past experience or known data [56]”

It is clear from the definition that Extrapolation is a technique in which past values are used to predict the future values. The relation of past values with other variables is also considered in this approach.

There are various methods of extrapolation available. Some of these are Linear, Polynomial, Conic, and French curve extrapolation. This method is basically a graphical method that fits a line or curve to the past value and predicts the future values according to the best fit line or curve [58].

However, the risks involved in these methods are that Cohorts are invisible in this method and current trend often do not continue [57].

2.9.4 Trend estimation

Trend estimation is a statistical tool which is used to set the trend of future values based on past observation. The trend estimation also has various models to predict the future values. These methods are similar to Extrapolation in the sense that these are also "lines of best fit" models. These models predict the values based on the best fit values according to the past values. However, these methods are not applied as different person will yield a different line of best fit. Some of the Trend estimation models are linear trend model, polynomial trend model, quadratic trend model, log linear trend model, and the exponential trend model

2.9.5 Growth curve

A growth curve is an empirical model of the evolution of a quantity over time. Growth curves are widely used in biology for quantities such as population size, body height or biomass. Values for the measured property can be plotted on a graph as a function of time [59]

Growth curves are employed in many disciplines besides biology, particularly in statistics, which has an extensive literature on growth curves. In mathematical statistics, growth curves are often modeled as beingcontinuousstochastic processes, e.g. as being sample paths that almost surely solve stochastic differential equations [60].

2.9.6 ARMA Models

ARMA stands for Auto Regressive Moving Average. This Method was proposed by Box and Jenkins in 1970 [63]. The First part, Auto Regressive means that previous values are used for forecasting the next value. Consider the time series with data from D1, D2, … Dn. The value D_{n+1} is predicted by considering the values up to D_n. Similarly the value D_{n+2} is predicted by taking the time series up to value D_{n+1}.

The second part of this model deals with Moving Average. Moving average is used to smooth the data. It eliminates cyclic, seasonal and irregular movements. However this method is sensitive to outliers therefore the outliers should be removed first from the time series or the sensitivity can be reduced by using weighted moving average [64].

ARMA models are designed to capture the way in which a stationary system moves back to its mean level after being pushed off it by a shock. The AR part models how previous values affect future ones whereas the MA part models how the shock itself affects future values.

Although both autoregressive and moving average approaches were already known, the contribution of Box and Jenkins was in developing a systematic methodology for identifying and estimating models that could incorporate both approaches. This makes Box-Jenkins models a powerful class of models. The next several sections will discuss these models in detail.

2.9.7 ARIMA Models

ARIMA is an acronym of Auto-Regressive Integrated Moving Average. This model is an extension of the ARMA model that incorporates trend. Trend can be linear or non-linear. Linear trend means the growth is based on a constant factor whereas non-linear means that the rate of growth changes over time too. This model is extensively used in the Time Series Analysis and forecasting.

A non-seasonal ARIMA model is classified as an ARIMA (p, d, q) model, where:

- p is the number of autoregressive terms,
- d is the number of non-seasonal differences, and
- q is the number of lagged forecast errors in the prediction equation.

2.10 Time Series analysis and Forecasting in EMIS

Since the field of EMIS is relatively new as compared to other fields of education therefore much work is needed in this field especially in the time series analysis and forecasting using EMIS database. Presently models are developed for forecasting but the projections of these models are not based on time series. Therefore there is a need of time series analysis and forecasting tool in the field of EMIS.

A tool, namely EMIS-Decision Support System (EDSS), is proposed in this study which uses the EMIS database as the major input. This database contains time series data about school education system. Since there were some catastrophe in the country like Floods in KP, Sindh and Baluchistan, terrorist activities in FATA, therefore the data of these provinces has not yet been finalized. The second input to the proposed tool is the population figures. The National Institute of Population Studies has been entrusted with the responsibility for providing official population statistics to the users. Therefore the population figures of NIPS are used in calculating various indicators. These figures are aggregated according to the age groups. For example age group 0-4 means that children who just born up to the age of four are aggregated in this group. Since various indicators require population of single age year, therefore it is difficult to disaggregate the age-group figures to singe age year figures. For this purpose, a tool namely Sprague Multiplier is used. This tool takes population figures according to age groups as input and distributes these figures to single age year. The details of this model can be seen in Appendix A.2.

When the input of the proposed tool has been set, the next step is to perform processing on these inputs. The processing is applied in two steps. The first step fetches the time series data from the database and applies the forecasting model to this data depending upon the disaggregation levels provided by the user. The ARIMA model is used as primary model for forecasting in this tool. However, another data mining forecasting method called Exponential Smoothing is also implemented. The user has the choice in using any of these two forecast models. The second step of processing is to calculate various EFA and MDG related indicators. These indicators are calculated up to the current year on EMIS data whereas the forecasted data is used to calculate the indicators beyond current year.

After the successful forecasting of variables and calculation of indicators, the result of these has to be presented to the users. The results are presented in the three forms. The first form presents the results in a spreadsheet like environment where all the Indicators related to the forecasted variables is shown. Various disaggregation levels provided by the user are also depicted. Since the graphical representation of data helps the user to understand the situation more easily therefore this data is also presented in the form of graphical line and bar charts. These charts help the education managers in quick understanding of education trends. The third format of presenting these results are in the form of GIS based maps. If the education managers are interested in seeing the individual institution's location, then this facility is provided by this tool. The education managers and policy makers can see the individual institution on the map.

The next chapters discuss these inputs, processing and outputs in detail.

In order to elaborate the overall architecture of the proposed tool, the following figure depicts the inputs, outputs and processing of the proposed tool.

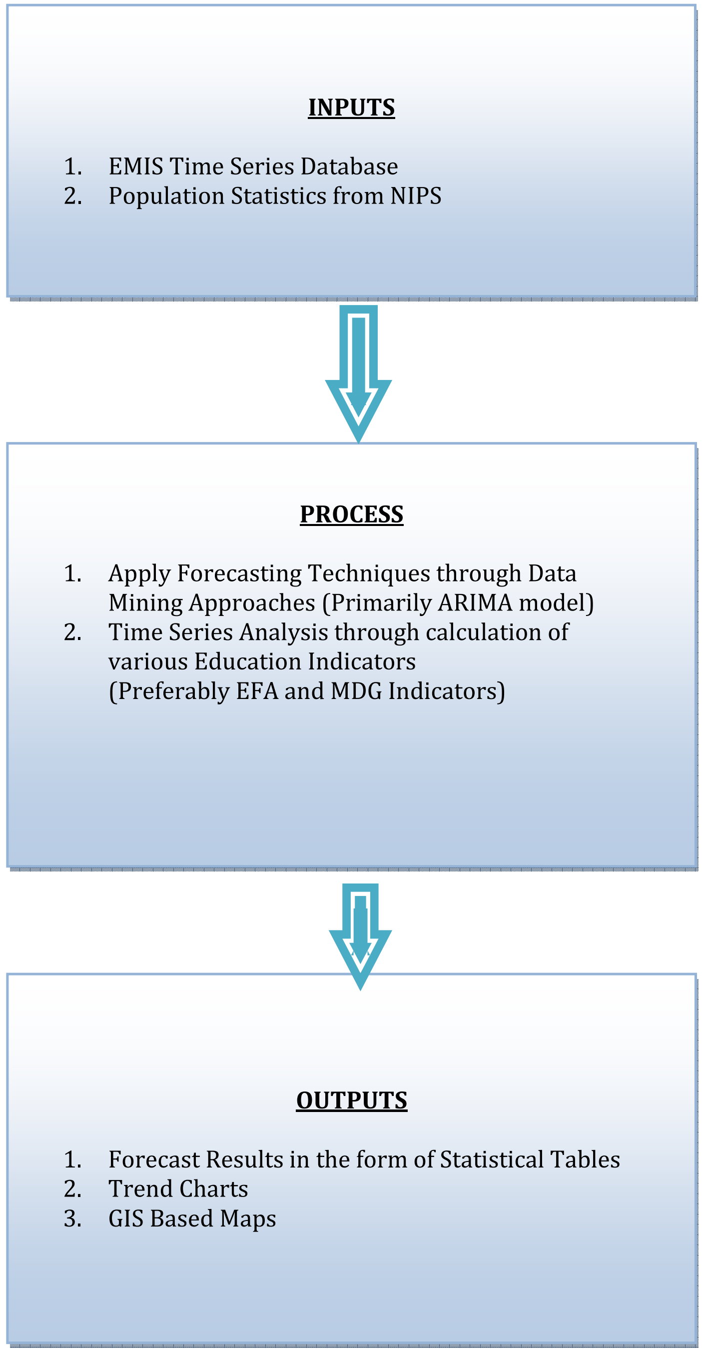


Figure 2.10:*Basic Flow chart of the proposed EDSS tool*

Chapter 3

DATA PREPARATION AND UNDERSTANDING

The significance of data cannot be denied in a decision making environment especially when the decisions are made about setting the directions in the future plans. Since this study provides a model for forecasting the education indicators using data mining techniques, therefore, the data needed for these forecasts and projections should be accurate and valid in order to get the reliable results.

The National EMIS is responsible for collating and consolidating the EMIS database at national level. This database contains the education data of schools for the past several years. This database is used as a primary source of data for the forecasts of education indicators in the present study. Before the National EMIS database is explained, it is important to understand the data collection process first.

3.1 Data Collection Process

The data is collected on the annual basis through the school census. It is a lengthy exercise which is done annually in order to get the latest education statistics. The date of 31st October of each is set by the government of Pakistan to conduct the Annual School Census. Therefore this date is used as reference date for school census. The annual school census consists of the following major steps:

1. Develop the Questionnaire
2. Survey Preparation and Questionnaire Distribution
3. Data collection
4. Data Entry
5. Data Consolidation at Provincial Level
6. Provincial Level Reports
7. Transfer and Consolidation of database at National EMIS
8. Report Generation, Indicators Calculation and Analysis at National EMIS

Here is the explanation of each of the above mentioned step:

3.1.1 Develop the Questionnaire

The first step of Annual School Census is the development of the questionnaire. The questionnaire is updated every year in order to get the information about the variables which are introduced as a result of latest changes in policy or trends in the education. Each province develops the province-specific questionnaire that contains the fields which needs to be filled by the head of institutions. Since the data of various variables is required at provincial and national level to calculate various indicators and get the latest data, therefore these questionnaires are prepared with due care. The development and finalization of questionnaire normally takes one month time. The questionnaire developed by Punjab EMIS for Annual School Census 2011-12 is placed at Appendix A.3.

3.1.2 Survey Preparation and Questionnaire Distribution

Once the questionnaire is developed and finalized, the next step is the distribution of questionnaire in the field. Before the distribution of questionnaire in the field, the data collectors are trained by the provincial EMIS. The training emphasizes on the collection of accurate and valid information.

Since the province is very large administrative, therefore the questionnaire s are distributed to a smaller administrative unit i.e. District. All the questionnaires of a district are sent to the Executive District Officer (Education). These questionnaires are then further distributed to a smaller unit i.e. Tehsil through Assistant District Officers (Education). The ADOs then distribute the questionnaire in the field with the help of school teachers. These teachers go from school to school and distribute the questionnaire.

The above mentioned activities of training and questionnaire distribution normally take two to three months.

3.1.3 Data collection

Each questionnaire is filled by the head of the institution. All the relevant information is extracted from the school record and provided in the questionnaire. The filled questionnaire is then returned back to the education authorities. This step normally takes two to three weeks as filled questionnaire collection from each school takes time.

3.1.4 Data Entry

When the questionnaires are filled by the head of institutions and returned back to the distribution authority then all these questionnaires are sent back to the provincial EMIS by adopting the same channel which is used to distribute the questionnaire i.e. from Tehsil Level to District Level to Province level.

After getting all the filled questionnaires, the data entry at provincial level is started. Some provinces perform data entry at district level whereas some perform this activity at provincial level.

The data entry duration is dependent on the number of questionnaires, length of questionnaires and number of individuals designated to data entry task. The data entry task normally takes two to three months.

3.1.5 Data Consolidation at the Provincial EMIS

In order to consolidate the entered data, the data validation and verification is performed. The data verification in some provinces is done by the third party whereas in some provinces it is done by the Monitoring and Supervision teams designated by the Education authorities. This activity spans over one month of time depending upon the verification and validation methodology.

3.1.6 Provincial Level Reports

Once the data is consolidated at provincial level, now it is ready to be published in the form of statistical report. The Provincial EMIS publish a statistical profile of the province, but these documents carry little information. This exercise takes ten to fifteen days to generate the provincial/district level reports.

3.1.7 Transfer and Consolidation of database at National EMIS

The generation of provincial reports represents the penultimate stage of the province's involvement in the EMIS cycle. The final role of the provinces is to transmit/transfer the consolidated education data for the province to NEMIS at the AEPAM. This represents the first instance of the movement of data in an exclusively electronic format. The database is converted from provincial level to national database structure. NEMIS constituted a committee called Technical Committee to perform data conversion from provincial database structure to national database structure. Technical Committee includes programmers from each province as well as from the NEMIS to perform the data conversion. The consolidated data is then verified by the provincial representatives.

3.1.8 Report Generation, Indicators Calculation and Analysis at National EMIS

After the finalization of national level database at NEMIS, the yearly publication namely "Pakistan Education Statistics" is then prepared and published. This publication is used by many national organizations involving education related activities as main source of data. The NEMIS is also responsible for providing the education statistics on demand basis.

On the basis of this publication, various education indicators are then calculated. The analysis is then performed about the current situation of education in the country. These indicators include Education for Indicators as well as Indicators included in the Millennium Development Goals.

The following picture elaborates the complete annual school census process:

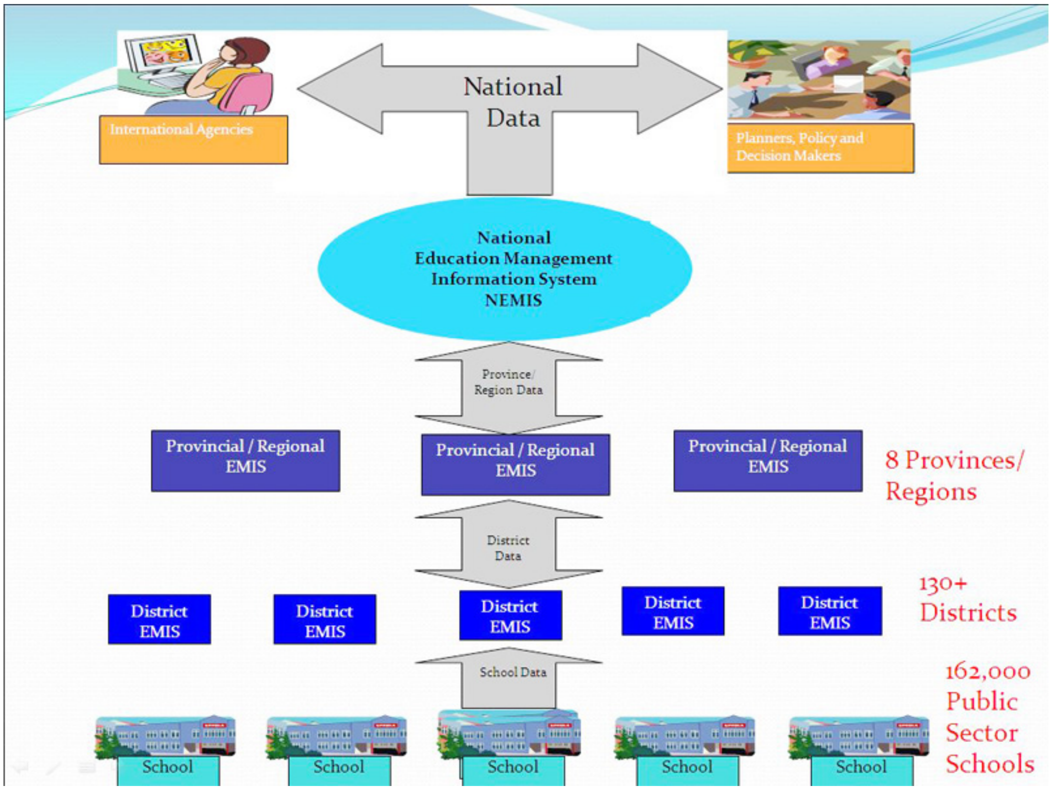


Figure 3.1:*Annual School Census Process*

3.2 Understanding the NEMIS Database Structure

The National EMIS database is managed in the Microsoft SQL Server database server. The database contains various tables that contain information regarding education system. Following are some of the important tables:

3.2.1 Institution

This table stores the basic information about the school.

3.2.2 Enrolment

The data related to students of the school is stored in this table disaggregated at the levels of class, gender and shift.

3.2.3 Building

This table keeps the information related to the building and physical facilities available in the school e.g. Class Rooms, Drinking Water etc.

3.2.4 Teachers by Academic Qualification

This table is used to store the information of the teachers with respect to their academic qualifications.

3.2.5 Teachers by Professional Qualification

Information of teachers about their professional qualification is stored in this table

3.2.6 Population

The information about the population disaggregated up to district level by gender and ager year is stored in this table.

3.2.7 Repeaters

Since various indicators require information about repetition of the enrolment in school, therefore it is pertinent to have the data about repeaters. This table is used to have the information about repeaters.

The following figure presents the overall diagram of NEMIS database:

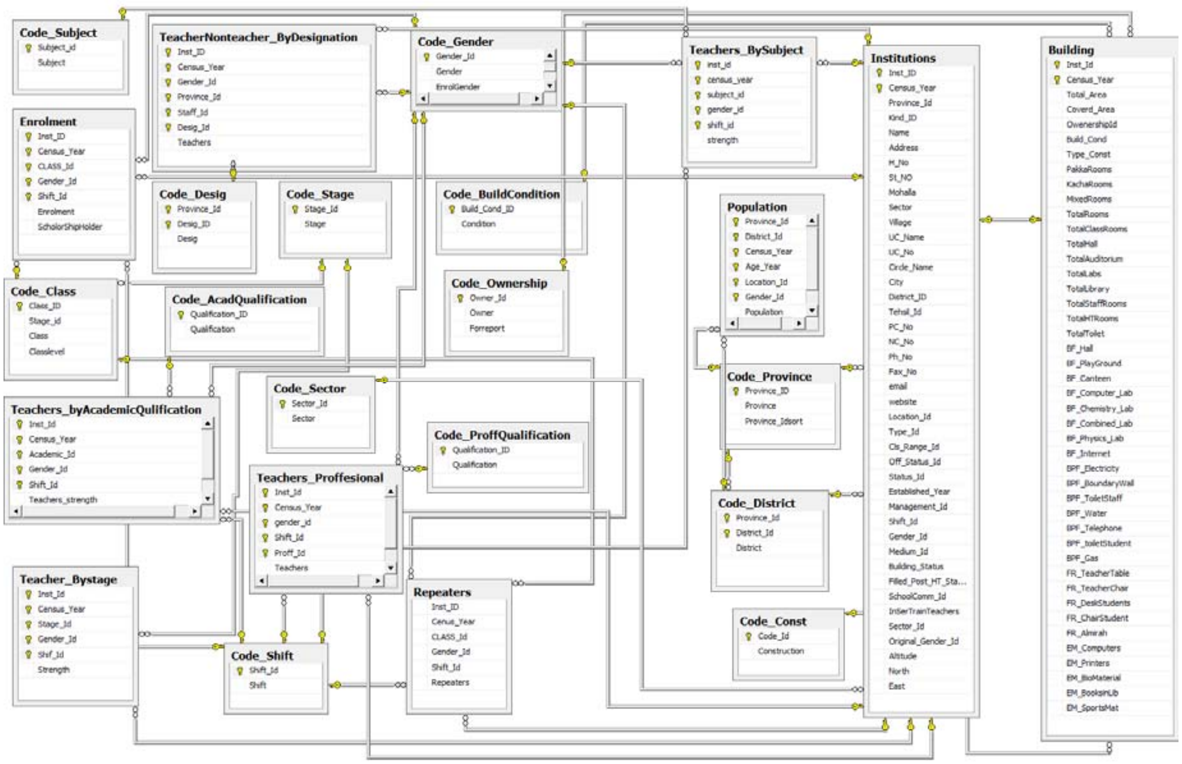


Figure 3.2:*Database Diagram of NEMIS database*

3.3 Understanding the NEMIS database variables

Each table of the NEMIS database contains various variables for which the information is stored. Following are the variables that are directly or indirectly used in the present study for calculation of indicators or forecasting the education statistics.

3.3.1 Variables Related to School and Its Building

a) Institution ID

Each school is assigned a unique identification number through the province. The field of Institution ID is used for this purpose.

b) Census Year

It is used to carry information about the census year for which the school data is entered.

c) District

It is used to identify the district in which the school is located.

d) Province

It is used to identify the province in which the school is located.

e) Tehsil

It is used to identify the tehsil in which the school is located.

f) Gender

The definition of gender varies from province to province. Some provinces determined the gender of school on the basis of enrolled students i.e. if only boys are studying in the school then the school gender is male. Similarly if girls are studying in the school then the school gender is determined as female. If both boys and girls are enrolled in the school, then the school gender is said to be mixed.

g) Location

This variable determines the location of school. Either the school will be located in rural locality or will be located in the urban locality.

h) Status

This field is used to store the information about the status of the school i.e. Either the school is functional, or the school is non-functional due to non-availability of students or teachers, or school is closed by the education authorities.

i) Type

The type of the institution determines the level of education which is instructed in the school e.g. Primary, Middle, High or Higher Secondary.

j) Shift

This variable stores the data about shifts which are operating in the school.

k) Medium

The medium of instructions is used to keep information about the languages which are used to teach the students.

l) Altitude / Longitude / Latitude

These variables are used to store information about the GIS coordinates of the institution.

m) Area

The information about overall area of the institution is stored in this variable.

n) Building Condition

This variable tells about the building condition i.e. Whether the Building is satisfactory or needs repair?

o) Building Ownership

Building Ownership variable is used to store data about the ownership of school building e.g. Whether the building is owned by government or the school building is donated by some person or organization or the school is presently operating in a rented building.

p) Class Rooms

The information about number of classrooms available in the schools is stored in this variable.

q) Toilet

This variable stores information about the availability of toilet in the school.

r) Boundary Wall

This variable stores information about the availability of boundary wall in the school.

s) Electricity

This variable stores information about the availability of electricity in the school.

t) Drinking Water

This variable stores information about the availability of drinking water in the school.

u) Playground

The information about availability of playground in the school is kept in this variable.

3.3.2 Variables Related to Teachers

a) Gender

It keeps the data of teacher's gender.

b) Designation

This variable contains the information about teacher's designation

c) Subject

The subjects which teacher is teaching is stored in this variable

d) Shift

The information of the shift in which the teacher is teaching is kept in this variable.

e) Basic Pay Scale

This variable is used to store information about the present Basic Pay Scale of the teacher.

f) Academic Qualification

This variable keeps data about academic qualification of degree e.g. Intermediate, Graduate, Post Graduate etc.

g) Professional Qualification

There are also various professional degrees available in the field of education. If any teacher possesses these qualifications, the information of such qualifications is stored in this variable.

h) Stage

This variable is used to store information about the stage of education to which teacher is teaching e.g. Primary, Middle, High etc.

3.3.3 Variables Related to Enrolment and Repeaters

a) Gender

It stores information about student's gender.

b) Class

This variable store the class in which student is studying e.g. Class 1 , Class 2 , Class 3 etc.

c) Shift

This variable is used to keep data about shift in which student is studying i.e. Morning or Evening

d) Scholarship Holder

This variable stores information about number of scholarship holder students.

3.3.4 Variables Related to Population

a) Age Year

This variable keeps data about the age year for which population is stored e.g. Year 4, Year 5 etc.

b) Population

This variable stores information about the population.

3.4 NEMIS Database Size

After having knowledge of NEMIS database structure as well as variables used to store various types of data, it is pertinent to know the number of records each table maintains. In the preceding paragraphs the number of records in each table is discussed.

The most important table in NEMIS database is Institution. All other tables are dependent on this table because the teachers' data, enrolment related data, and information related to repeaters will not be stored until and unless the institutions data is available in the Institutions table. Approximately this table contains 168,000 records for each year. If NEMIS database contains data for 10 years then there will be approximately 1.6 million records in this table. Among these 165,000 records, Approximately 65000 records belongs to the province of

Punjab, 45,000 to Sindh, 12000to Baluchistan, 33,000 to KP, 400 to ICT, 5500 to AJK, 5500 to FATA and 1500 to Gilgit-Baltistan.

The table containing data about Enrolment in a school is second most important table. Since the enrolment is maintained in this table at various disaggregated levels, therefore, the records in this table are much more than the Institutions table.For a single year, this table contains 1.5 million rows of enrolment related data. Out of these, 850000 rows contains enrolment data of Punjab, 70000 of Baluchistan, 190000 of KP, 290000 of Sindh, 3000 of ICT, 19000 of GB, 29000 of FATA, and 49000 of AJK. Therefore, if NEMIS database contains data of a decade, then this table will be maintaining approximately 15 million records. Since this table contains a huge number of records, therefore, various types of clustered and non-clustered indexes are used in order to fetch the data from database in the least possible time.

The education system is incomplete without teachers. Therefore, the teachers' data is also maintained in NEMIS database.Teachers' data is maintained according to their gender, academic, and professional qualifications. The teachers' data disaggregated by Academic qualification is maintained in the Teachers by Academic Qualification table whereas the teachers' data disaggregated by Professional Qualification is stored in Teachers by Professional Qualification table. Professional Qualification means the degree or certificate obtained for professional teaching at various levels of education.

The Teachers by Academic Qualification table contains 310000 rows for a single year data. The teachers' data of Punjab Province is stored in 134000 records. Similarly 70000 records are used to store data of Sindh, 56000 for KP, 19000 for Baluchistan, 1500 for ICT, 4500 for GB, 12000 for FATA and 13000 for AJK.

There are approximately 300000 records in Teachers by Professional Qualification for a single year. 136000 records contains teachers data of Punjab, 66000 of Sindh, 51000 of KP, 18000 of Baluchistan, 1500 of ICT, 13000 of AJK, 10000 of FATA, and 4500 of GB.

The data pertaining to repeaters is also very important because the repeaters data is used to calculate various indicators related to internal efficiency of the education system. Therefore, the repeaters table is designed to store the information related to repeaters. The repeaters data is disaggregated to the same levels as of enrolment. However this table contains much less rows than enrolment. This table approximately has 300000 rows for single year data. Out of these 300000 records, 87000 are of Punjab, 80000 of Sindh, 60000 KP, 25000 Baluchistan, 2000 of ICT, 5000 of GB, 31000 of FATA and 7000 of AJK.

Since the data related to classrooms are used to calculate Pupil-Classroom Ratio, therefore it is very important to store the information related to physical infrastructure available in the building. For this purpose the table of Building is used. This table contains same number of rows as of Institutions table.

The Population is used to calculate various indicators related to coverage of the education system e.g. Gross Enrolment Ratio, Net Enrolment Ratio. Therefore the Population table is used to store information related to population disaggregated by Province, District, Age, Gender, Location and Year. The Government of Pakistan has authorized National Institute of Population Studies (NIPS) as official population figure provider. Therefore NEMIS obtain the population figures from NIPS and populate these in the NEMIS database.

Chapter 4

IMPLEMENTATION AND RESULTS

Since the data is processed and prepared to be used as input for the time series analysis and forecasting, now the next step is to apply the time series model to the data set. Microsoft Visual C# 2012 is used to implement the time series models on the data set. Graphical User Interface has been designed to capture the user input. Various classes have been written to apply the time series model on the data set depending upon the user input. The proposed tool will be referred as EMIS-Decision Support System (EDSS) throughout this thesis.

4.1 User Interface Design

Following picture shows the main screen of EDSS

Figure 4.1:*Startup Screen of EDSS*

The EDSS provides following three major functionalities to the Education Managers:

1. Forecasting
2. Comparisons
3. Top-N Analysis (conversely Bottom N Analysis)

Therefore, the user interface for each of the above functionality is required. The following sections discuss the user interfaces designed to capture the input options required for above mentioned functionalities.

4.1.1 Forecasting

This option enables the Education Managers to forecast the variable. The education manager can disaggregated the forecast at levels of District, Location, Level, Gender and Sector. Furthermore the manager can also set the number of years to be forecasted.

The following figure shows the form which is designed to capture the Base year and end options from the user:

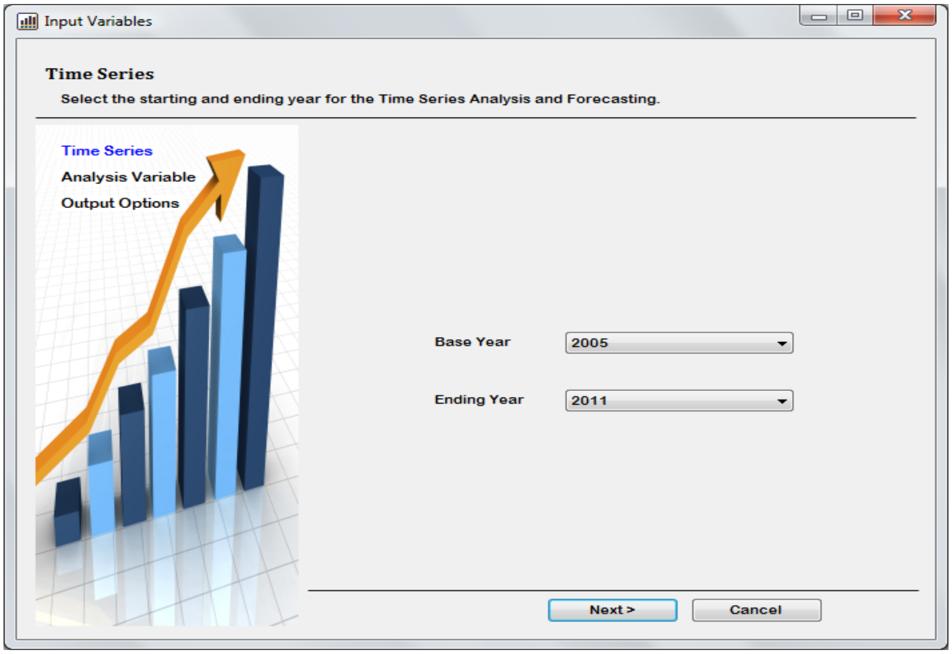


Figure 4.2: *User Interface to capture Time Series observations*

The base year is used to provide the starting point for observations. If base year is set to 2005 then it means that the observations from year 2005 to onwards will be used for forecasting the future values. Similarly, the ending year is required to complete the set of observations. If the ending year is set to 2010 then it means that the observations upto year 2010 will be used for forecasting. This option is very important when dealing with large datasets. For instance, if the database contains data for 40 years then there is a need to specify the number of observations that will be used in forecasting process.

When the number of observations are provided/selected then the next step is to provide various disaggregation levels for which forecasting is required. In order to capture the disaggregation levels, another form is designed. This form contains input fields which users can set in order to disaggregate the data set. Following is the screenshot of this form:

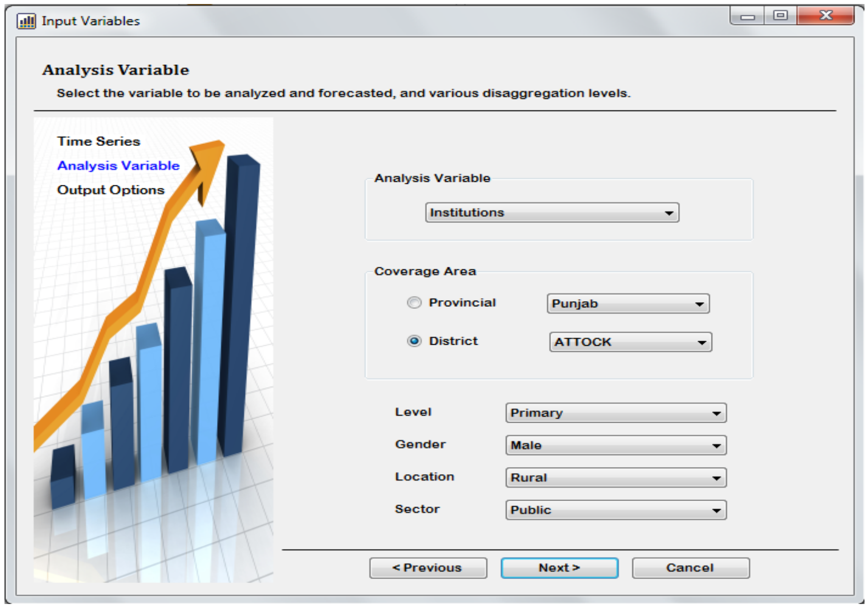


Figure 4.3:*Input Screen to record Analysis Variable and various disaggregation levels*

The first option of this form is Analysis Variable. Since the NEMIS database contains various variables which can be forecasted, therefore, it is imperative to select a single variable to be forecasted. The next disaggregation level is the Province/District. The NEMIS database contains school's data from all over the Pakistan so it is very necessary to analyze the performance of a single Province/District. For this purpose the users are required to select the desired Province and/or District for which forecasting is required. Another important disaggregation is level of education. This aggregation level is used most of the time as the policy and decision makers require data for a specific level of education to focus on. The disaggregation of variable by gender is also very important when analyzing gender discrimination and disparities. This disaggregation further helps in focusing the discriminated/dispirited areas. The location variable defines the locality of the school in the sense that whether the school is located in urban area or rural area. Since two-third of Pakistan's population is residing in the rural areas, therefore, it is very important to analyze the access of education and state/performance of education system in these areas. The last disaggregation level is the sector e.g. public or private. If the school is function under the administrative control of Department of Education then it is termed as public sector school whereas if the school is functioning autonomously without intervention of any government department then it is said to be private sector school. This disaggregation is very important when we are allocating government resources for the schools. In view of these, the disaggregationis must in analyzing and forecasting the education statistics.

When the disaggregation levels are set, then it is time to choose the forecasting model and number of years for which the forecasting is required. Two forecast models namely “ARIMA” and “Exponential Smoothing” are implemented in this study therefore the users are provided with the option to choose anyone as their forecasting model. The next option is to set the number of years for which the forecasted data is required. The minimum number of forecasts step is 1. As the graphs and charts depict the information in visual style and increase understanding, therefore, the option of charts is also provided. The users can choose the values for which the graph is required e.g. Graph of actual values, graph of forecasted values or graph of both actual and forecasted values. The form designed to capture these input options is shown here:

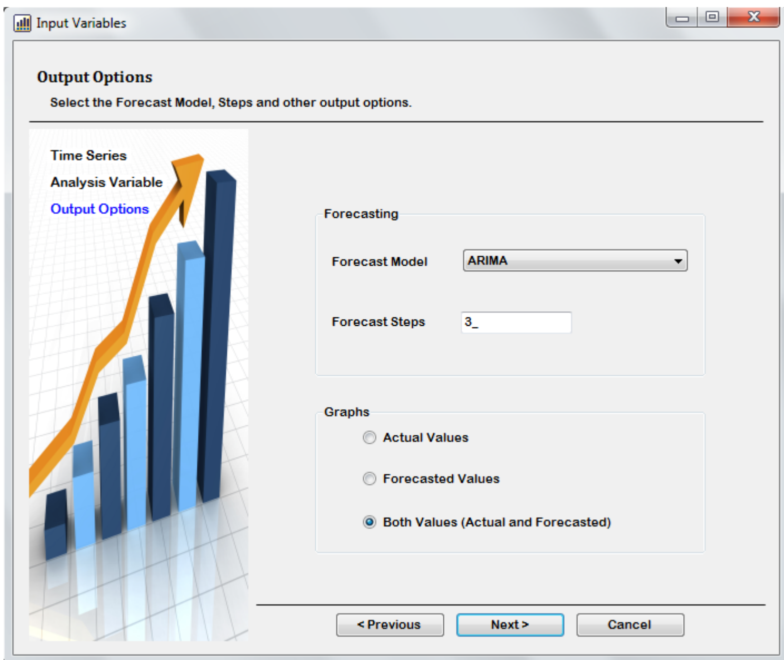
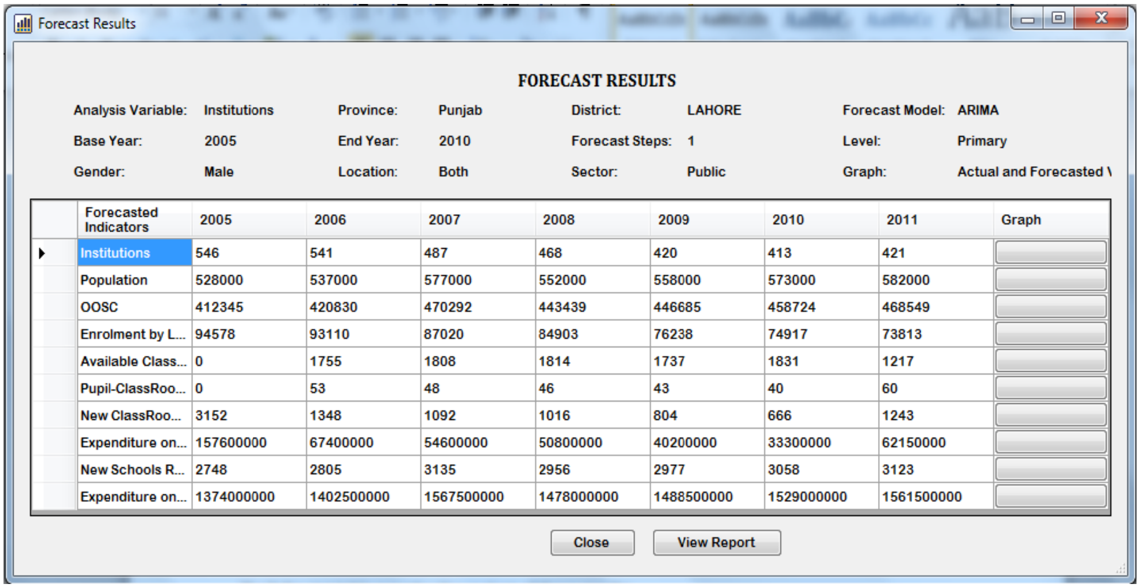


Figure 4.4: *User Interface for selection of forecasting model and steps*

When all of the required parameters are provided to the tool, the next step is to perform the forecast and display the result. The following screen shows the forecasts results produced by the tool:



Forecast Results

FORECAST RESULTS

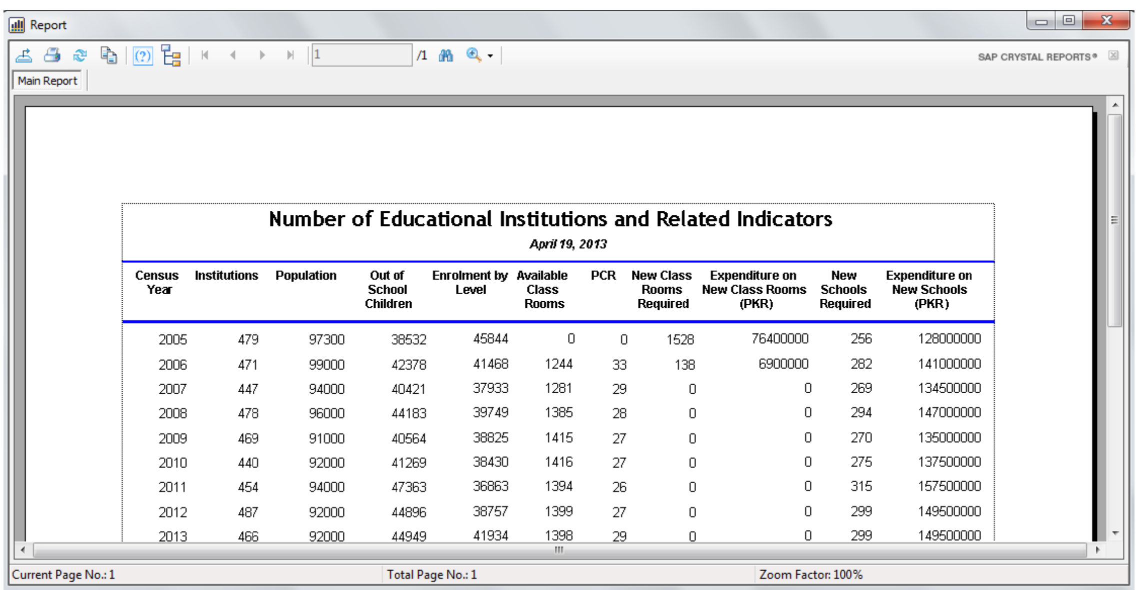
Analysis Variable: Institutions | Province: Punjab | District: LAHORE | Forecast Model: ARIMA
Base Year: 2005 | End Year: 2010 | Forecast Steps: 1 | Level: Primary
Gender: Male | Location: Both | Sector: Public | Graph: Actual and Forecasted \

Forecasted Indicators	2005	2006	2007	2008	2009	2010	2011	Graph
Institutions	546	541	487	468	420	413	421	
Population	528000	537000	577000	552000	558000	573000	582000	
OOSC	412345	420830	470292	443439	446685	458724	468549	
Enrolment by L...	94578	93110	87020	84903	76238	74917	73813	
Available Class...	0	1755	1808	1814	1737	1831	1217	
Pupil-ClassRoo...	0	53	48	46	43	40	60	
New ClassRoo...	3152	1348	1092	1016	804	666	1243	
Expenditure on...	157600000	67400000	54600000	50800000	40200000	33300000	62150000	
New Schools R...	2748	2805	3135	2956	2977	3058	3123	
Expenditure on...	1374000000	1402500000	1567500000	1478000000	1488500000	1529000000	1561500000	

Close | View Report

Figure 4.5:*User Interface to display forecasts results*

The top portion of this form displays the values of disaggregation selected by the user. The forecasted variable and the indicators related to these variables are displayed in tabular form in the lower part of the screen.



Report

SAP CRYSTAL REPORTS®

Main Report

Number of Educational Institutions and Related Indicators

April 19, 2013

Census Year	Institutions	Population	Out of School Children	Enrolment by Level	Available Class Rooms	PCR	New Class Rooms Required	Expenditure on New Class Rooms (PKR)	New Schools Required	Expenditure on New Schools (PKR)
2005	479	97300	38532	45844	0	0	1528	76400000	256	128000000
2006	471	99000	42378	41468	1244	33	138	6900000	282	141000000
2007	447	94000	40421	37933	1281	29	0	0	269	134500000
2008	478	96000	44183	39749	1385	28	0	0	294	147000000
2009	469	91000	40564	38825	1415	27	0	0	270	135000000
2010	440	92000	41269	38430	1416	27	0	0	275	137500000
2011	454	94000	47363	36863	1394	26	0	0	315	157500000
2012	487	92000	44896	38757	1399	27	0	0	299	149500000
2013	466	92000	44949	41934	1398	29	0	0	299	149500000

Current Page No.: 1 | Total Page No.: 1 | Zoom Factor: 100%

Figure 4.6:*Report generated by EDSS tool*

Since the NEMIS also has the responsibility of disseminationof data and statistics, therefore, the forecast results should be available in a report form. For this purpose, the View Report button is available in the bottom of Screen. When the user clicks on this button, the report containing the forecasts result is generated. The figure 4.6 shows the report generated by EDSS tool.

At the end of each indicator the graph option is provided. If the user wants to view the graph of the indicator then he simply needs to click the graph button. The figure 4.5 shows the graph generated by EDSS tool. The number of observations used to carry out the forecasting is presented in Green color whereas the forecasted values are displayed in blue color. The horizontal axis shows the years whereas the vertical axis displays the data for selected indicator. The line charts are very important in easy understanding of Trends.

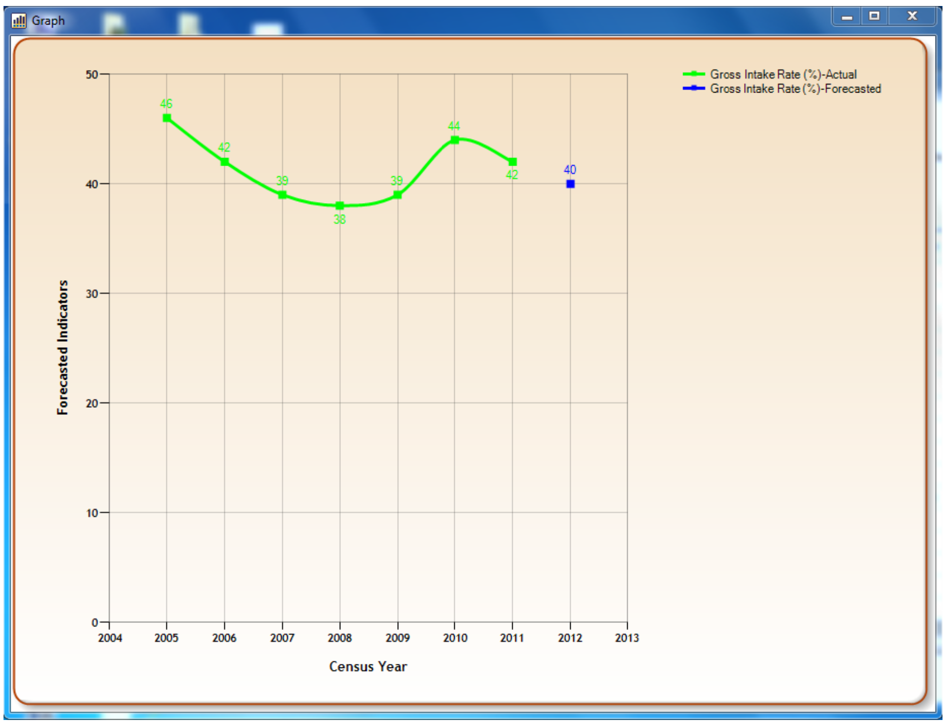


Figure 4.7:*Graph produced by EDSS tool showing actual and forecasted values*

4.1.2 Comparisons

It is often required in educational planning to compare two or more districts or provinces to compare their performance. In order to implement this flexibility, the following forms have been designed which capture the input of the user to compare two or more districts/ provinces:

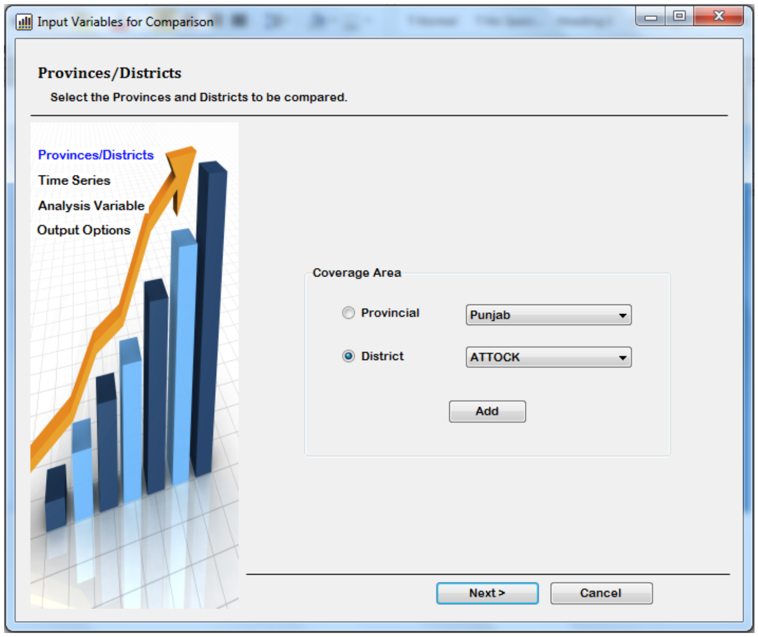


Figure 4.8:*User interface for selection of districts/provinces for comparison*

In order to compare the performance/variables of two or more districts or provinces, the provinces/districts are required to be selected first. The figure 4.7 shows the input screen used to capture the districts and provinces for comparison. The user simply needs to select the desired province and district from the drop down list and then click add. The user will repeat this process until all the desired districts/provinces have been added for comparison. After it, the user needs to click the Next button. All the other options related to various disaggregation levels, forecast models, steps, and graphs are same as of forecasting the single variable.

4.1.3 Top-N Analysis

The significance of top-n analysis cannot be denied as it helps us in assessing the performance of any number of top/bottom districts or provinces. The following user interface has been designed to capture user inputs for these purposes:

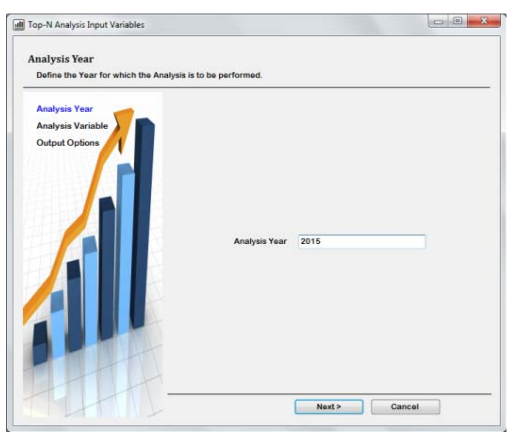


Figure 4.9:*Input Screen to capture the Top-N Analysis year*

The analysis year is very important to capture when performing Top-N Analysis because the ranking of districts and provinces are calculated for a specific year. All the other options related to various disaggregation levels, forecast models, steps, and graphs are same as of forecasting the single variable.

4.2 Implementation Classes

Once the input options have been captured from the user, the next step is to perform time series analysis and forecasting on the data. Since the tool is developed in the Microsoft Visual C# therefore classes has been written to perform time series analysis and forecasting. Following is brief of each implementation class

4.2.1 clsDatabase

This class is used to get data from the database. Since we are using EMIS database therefore various information is required from the database e.g. enrolment, teachers, institutions, and population. In order to get this information from database, this class is implemented to provide connection to database. This class contains various functions which fetch the information from database on the provided parameters. Here is the overall structure of this class:

```
classclsDatabase
{
public clsDatabase()
publicbool OpenConnection()
publicvoid CloseConnection()
publicDataSet GetYears()
publicDataSet GetProvinceDistrict(string district)
publicDataSet GetAllDistricts(int province)
publicDataSet GetAllDistrictsPunjab()
public System.Data.DataTable GetTimeSeries(string variable)
public System.Data.DataTable GetGISCoordinates()
public System.Data.DataTable GetPopulation()
public System.Data.DataTable GetPopulationbyAgeYear(int Age_Year)
public System.Data.DataTable GetClassRooms()
public System.Data.DataTable GetEnrolmentByLevel()
public System.Data.DataTable GetEnrolmentbyClass(int Class)
public System.Data.DataTable GetRepeatersbyClass(int Class)
}
```

4.2.2 clsForecast

Once the time series has been fetched from the database, the next step is to perform forecasting. clsForecast contains various functions which performs forecasting. Following lines presents the overall structure of this class and functions:

```
classclsForecast
{
	publicDataTable ForecastTimeSeries(string Forecast_Model, DataTable dtb, string title)
publicDataTable GetInstitutionsForecastedSeries()
publicDataTable GetEnrolmentForecastedSeries()
publicDataTable GetEnrolmentbyLevelForecastedSeries()
publicDataTable GetAvailableClassRoomsForecastedSeries()
publicDataTable ForecastInstitutions()
publicDataTable ForecastEnrolment()
publicDataTable ForecastTeachers()
publicDataTable ForecastPhysicalFacilities()
publicDataTable GetForecastedClassWiseEnrolment(uint level)
publicDataTable GetForecastedClassWiseRepeaters(uint level)
}
```

4.2.3 clsIndicators

When the forecasting has been performed, the next step is to calculate various indicators related to selected variables and various disaggregation levels. A separate class namely clsIndicators has been implemented to calculate education indicators. Following is the structure of this class.

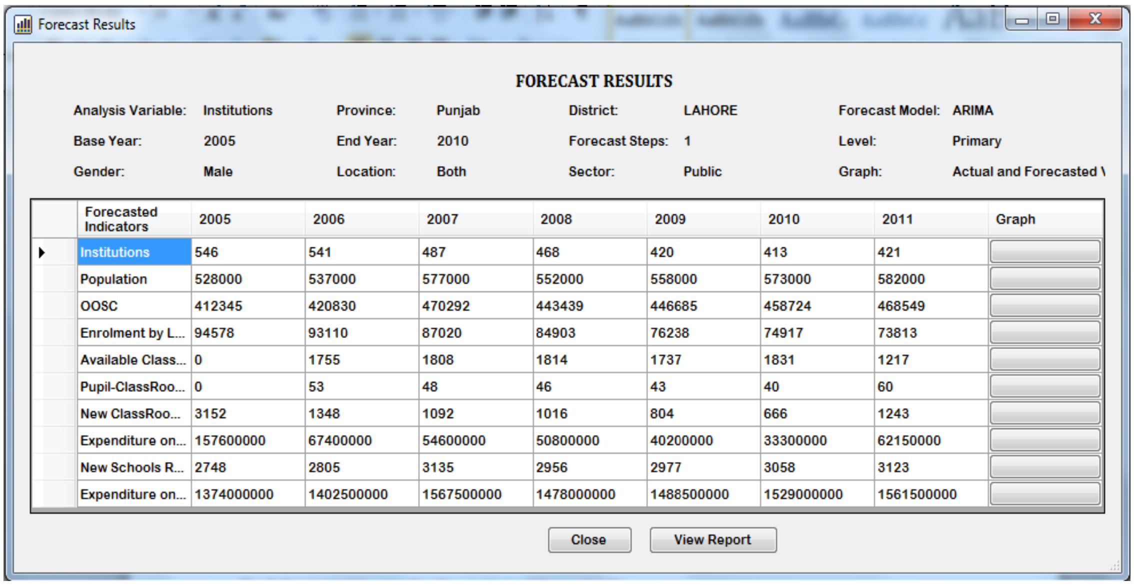
```
classclsIndicators
{
publicDataTable GetOOSC(DataTable Enrolment, DataTable Population)
publicDataTable GetPCR(DataTable EnrolmentbyLevel, DataTable ClassRooms)
publicDataTable GetNewRoomsRequired(DataTable EnrolmentbyLevel, DataTable
 ClassRooms)
publicDataTable GetCostofNewRooms(DataTable ClassRooms)
publicDataTable GetNewInstitutionsRequired(DataTable OutofSchoolChildren)
publicDataTable GetCostofNewInstitutions(DataTable Institutions)
publicDataTable GetGER(DataTable Enrolment, DataTable Population)
publicDataTable GetNER(DataTable Enrolment, DataTable Population)
publicDataTable GetRepeaters(DataTable ClassWiseRepeaters)
		publicDataTable GetPercentageofRepeaters(DataTable Enrolment, DataTable
		Repeaters)
publicDataTable GetGIR(DataTable ClassWiseEnrolment, DataTable
	ClassWiseRepeaters, DataTable Population)
publicDataTable GetNIR(DataTable ClassWiseEnrolment, DataTable
	ClassWiseRepeaters, DataTable Population)
publicDataTable GetTransitionRate(DataTable ClassWiseEnrolment, DataTable
	NextClassEnrolment, DataTable NextClassRepeaters, uint level)
publicDataTable GetPTR(uint level)
publicDataTable GetTeachersRequired(DataTable EnrolmentbyLevel, DataTable
	AvailableTeachers)
publicDataTable GetExpenditureofTeachersRequired(DataTable Teachers)
publicDataTable GetPercentageofInstitutionswithFacility(DataTable Institutions,
DataTable InstitutionswithFacility)
publicDataTable GetExpenditureonFacility(DataTable Institutions, DataTable
	InstitutionswithFacility, long expense)
public System.Data.DataTable CalculateSurvivalRate(System.Data.DataTable
	enrolment, System.Data.DataTable repeaters)
}
```

4.3 Results

As the Data Set was prepared and the tool was designed to implement the forecasting model on the dataset, now it is time to generate the results and perform analysis.

4.3.1 Forecast Results

The following figure shows the result of Institution variable forecasted by the EDSS for the year 2011.



Forecasted Indicators	2005	2006	2007	2008	2009	2010	2011	Graph
Institutions	546	541	487	468	420	413	421	
Population	528000	537000	577000	552000	558000	573000	582000	
OOSC	412345	420830	470292	443439	446685	458724	468549	
Enrolment by L...	94578	93110	87020	84903	76238	74917	73813	
Available Class...	0	1755	1808	1814	1737	1831	1217	
Pupil-ClassRoo...	0	53	48	46	43	40	60	
New ClassRoo...	3152	1348	1092	1016	804	666	1243	
Expenditure on...	157600000	67400000	54600000	50800000	40200000	33300000	62150000	
New Schools R...	2748	2805	3135	2956	2977	3058	3123	
Expenditure on...	1374000000	1402500000	1567500000	1478000000	1488500000	1529000000	1561500000	

Figure 4.10: *Forecast Results for Institutions variable*

When the interface design has been completed, and the implementation classes have been written and implemented, the next step is to carry out the forecasts and analyze the results. First of all, the institutions variable is analyzed. The forecasts for the year 2011 were made for Public Sector Male Institutions of Primary level located in Lahore District. The base year was set to 2005 whereas 2010 was the ending year. This forecast was done by using ARIMA as well as Exponential Smoothing model. The following table provides the analysis of these results:

Table 4.1: *Illustration of forecasted results of Institutions variable*

Institutions			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	420	421	1
Exponential Smoothing	420	413	7
Trend Line	420	392	28
Moving Average	420	417	3

It is evident from the above table that the ARIMA model predicted the value of 421 whereas the actual value is 420. Similarly Exponential Smoothing model predicted the value of 413 for year 2011. The conventional growth rate method predicted the value to be 392. This shows that the ARIMA model predicted the closest value among these three models.
The reason for best performance of ARIMA model in this case is that the time series given to it has decreasing trend. The value of base year was 546 which continues until 413 in 2010. The ARIMA model most appropriately calculated the value of decreasing trend in this time series. This decreasing trend value was further used in forecasting the next year's result.
When the Institutions variable is forecasted, the indicators related to this variable are also forecasted. One of these indicators is Out of school children. The following table shows the predicted as well as actual values of Out of School Children for the year 2011.

Table 4.2:*Illustration of forecasted results of Out of School Children Indicator*

Out of School Children			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	466,536	468,549	2,013
Exponential Smoothing	466,536	469,000	2,464
Trend Line	466,536	469,043	2,507
Moving Average	466,536	452,704	13,832

It can clearly be seen from the above table that the closest value is predicted by ARIMA model. The ARIMA model predicted that in the year 2011 there will be 468,549 children out of school for the given parameters and disaggregation against the actual value of 466,536. The Exponential Smoothing model predicted the second closer value of 469,000.
The time series data of Out of School Children was given to the ARIMA model. This time series was having some fluctuations in value due to the fact that this indicator depends on the values of population. The ARIMA model, as compared to other models, more appropriately measured the variations in the time series. The forecast of next year was made by considering this fluctuation in the time series. That is why the prediction made by ARIMA model is closest to the actual value.
Enrolment by Level is another indicator which is calculated along-with Institutions Variable. The following table shows the predicted and actual values of Enrolment by Level for the year 2011.

Table 4.3: *Forecast Results of Enrolment by Level Indicator*

Enrolment by Level			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	74,635	73,813	822
Exponential Smoothing	74,635	74,917	282
Trend Line	74,635	71,802	2,833
Moving Average	74,635	75,578	943

The above table shows that the Enrolment by Level indicator is most closely predicted by Exponential Smoothing model with a difference of only 282. Trend Line, same like previous indicators, produced very poor results.

The Pupil-Classroom Ratio Indicator was forecasted next. The following table shows the result of forecast for single year.

Table 4.4: *Forecast Results of Pupil-Classroom Ratio Indicator*

Pupil-Classroom Ratio			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	42	43	1
Exponential Smoothing	42	42	0
Trend Line	42	38	2
Moving Average	42	41	1

The table shows that the Exponential Smoothing method produced the exact results. The ARIMA model also produced good results with a difference of only 1. But the traditional Trend Line model produced poor results.

In the above two indicators, namely Enrolment by Level and Pupil-Classroom Ratio, the performance of Exponential Smoothing model is excellent. In the case of Pupil-Classroom Ratio, it predicted exact results as of actual whereas in Enrolment by Level, it produced the nearest match. The reason behind this is that the time series provided to this model has continuous decreasing/increasing trend. Since exponential smoothing model give more weight to last year value and lesser weight to the base year values, therefore, the prediction heavily depends on last two to three years values. This forms the basis for exponential smoothing method to produce closest predictions.

It is evident from the above tables that the best forecast values are predicted by either ARIMA model or Exponential Smoothing Model. After successfully forecasting the variables for one year, the values for two or more years were forecasted. Following tables shows the results of the forecasts for two years:

Table 4.5: *Illustration of two years forecast results of Out of School Children Indicator*

Institutions					
Model	**Actual Value**		**Predicted Value**		**Mean Absolute Error (MAE)**
	2010	**2011**	**2010**	**2011**	
ARIMA	413	420	404	427	8.0
Exponential Smoothing	413	420	420	420	3.5
Trend Line	413	420	396	374	31.5
Moving Average	413	420	444	432	21.5

According to the above table, when the two years forecasts were made, the Exponential Smoothing performed very well as compared to other two models. The Mean Absolute Error is smallest i.e. 3.5 which shows a minor error in prediction.

Since the exponential smoothing method gives more weight to the last two to three years values, therefore, the prediction made by this method is mostly dependent on these values. This is the major reason that the exponential smoothing method performed excellently in two years forecasts as compared to other models which give equal weight to each observation in calculating the trend/variations in the time series.

Table 4.6: *Illustration of two year forecast results of Out of School Children Indicator*

Out of School Children					
Model	**Actual Value**		**Predicted Value**		**Mean Absolute Error (MAE)**
	2010	**2011**	**2010**	**2011**	
ARIMA	458,724	466,536	458,278	469,286	1598.0
Exponential Smoothing	458,724	466,536	462,461	470,940	4070.5
Trend Line	458,724	466,536	455,985	465,637	1819.0
Moving Average	458,724	466,536	445,062	445,874	17162

When the Out of school children indicator was analyzed for two years, it was found that ARIMA model performed better with minimum Mean Absolute Error among these models. ARIMA model predicted that there will be 458,278 out of school children in 2010 and 469,286 in 2011. So it enables the policy and decision makers to focus these children and provide access of education to these children.

When we predicted the value of Out of School Children Indicator for a single year, It was found that the ARIMA model more appropriately measured the variations in the time series. Now we have predicted the value of this indicator for two years, it is found that the ARIMA model produced the expected results and predicted the closest values for two years. The major reason behind this is that the ARIMA model more appropriately measured the fluctuations in the time series.

Similarly the Enrolment by Level indicators was predicted for two years. The following table shows the forecasts results.

Table 4.7:*Two years Forecast Results of Enrolment by Level Indicator*

Enrolment by Level					
Model	**Actual Value**		**Predicted Value**		**Mean Absolute Error (MAE)**
	2010	**2011**	**2010**	**2011**	
ARIMA	74,917	74,635	94,578	94,578	19802
Exponential Smoothing	74,917	74,635	76,238	76,238	1462
Trend Line	74,917	74,635	72,542	69,162	3924
Moving Average	74,917	74,635	80,571	78,404	4711.5

This table shows that the Exponential Smoothing model has best predicted the value of 76,238 children enrolled in primary level of education in 2010 and 2011.

The tables 4.4 to 4.6 shows that the ARIMA and Exponential Smoothing techniques showed the closest forecasts as depicted by Performance Measures.

After the successful forecasting of Institutions variable, the next variable Enrolment was forecasted. The forecast was done for the Enrolment of Public Sector Middle Schools of Faisalabad Districts located in the rural areas. The base year was set to 2005 whereas 2010 was the ending year. Following figure shows the forecasted results produced by EDSS.

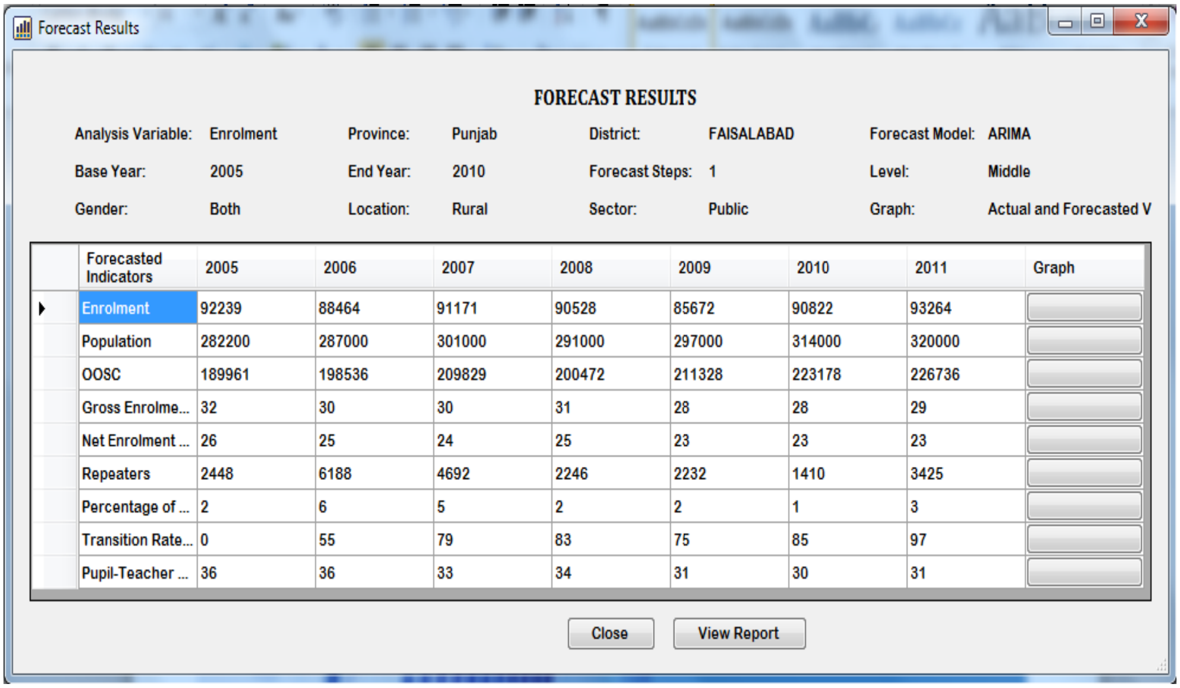


Figure 4.11:*Forecast Results for Enrolment variable*

Following table presents the analysis of the above mentioned forecast results.

Table 4.8:*Illustration of forecast results of Enrolment variable*

Enrolment			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	93,891	93,264	627
Exponential Smoothing	93,891	89,952	3939
Trend Line	93,891	90,543	3348
Moving Average	93,891	88,247	5644

The ARIMA model predicted that there will be 93,264 students enrolled in the middle stage in year 2011 whereas the actual value is 93,891. Hence the difference in absolute number is 627 which is negligible. The Trend Line method produced the second closer prediction with value of 3348 students enrolled in the middle stage.

The value of enrolment variable was predicted. The time series used for this purpose contains values from year 2005 to 2010. This time series do not contain any decreasing or increasing trend rather it contains fluctuations. In this case, the ARIMA model produced better results due to the fact that it has the capability to properly calculate the fluctuations in the time series. The reason behind poor performance of Trend Line method is that it only produces exact results in the cases where time series have increasing/decreasing trend.

Like the institutions variable, when the enrolment variable is forecasted various other indicators are also calculated. Repeaters are one of these indicators. The following table shows the forecasts results of Repeaters indicator:

Table 4.9:*Illustration of forecast results of Repeaters Indicator*

Repeaters			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	2,448	3,425	977
Exponential Smoothing	2,448	2,476	28
Trend Line	2,448	1,291	1157
Moving Average	2,448	1,821	627

It can be seen that the Exponential smoothing method produced the closest value with a difference of only 28. The ARIMA model predicted the second closer value of 3425 against the actual value of 2448. Like the institutions variable and related indicators, the Trend Line method produced the poor results.

The major reason behind good results by Exponential Smoothing method is that the time series of repeaters do not have much variations in the data as well as there is not any increasing/decreasing trend. The Exponential Smoothing method work very well in the absence of variations and trend. Therefore, the exponential smoothing method produced close results to actual value.

Transition Rate is another indicator which is used to analyze that how many students are enrolled in the next grade that completed the previous grade of education. Ideally this value should be 100%. However due to various reasons it never reached 100% in Pakistan. The following table shows the comparison of values of Transition Rate indicator:

Table 4.10:*Forecast results of Transition Rate Indicator*

Transition Rate			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	88	97	9
Exponential Smoothing	88	90	2

Trend Line	88	97	9
Moving Average	88	80	8

It is evident from the above table that the closest value of 90 is predicted by the Exponential Smoothing Method while ARIMA produced second closer results.

The Transition Rate indicator is calculated by using the enrolment values of consecutive two years. It means that the forecast results will be dependent on two years. In this case the Exponential Smoothing method is most appropriate as it gives more weightage to the previous values. That is why it produced closest results in the Transition Rate indicator.

The Gross Enrolment Ratio indicator is used to analyze the access of education to the population. In other words it measures the coverage of education system. However this indicator considers students irrespective of their age group. The following table shows the results of Gross Enrolment Ratio Indicator:

Table 4.11:*Forecast results of Gross Enrolment Ratio Indicator*

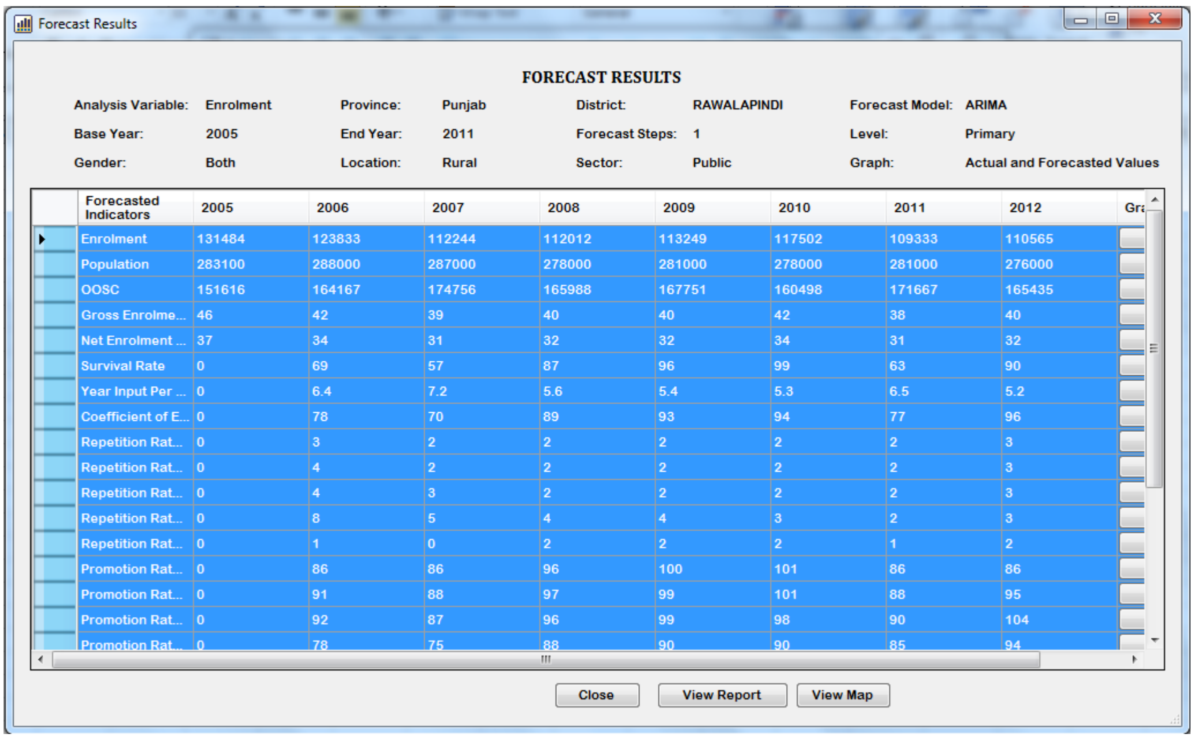
Gross Enrolment Ratio			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	28	29	1
Exponential Smoothing	28	28	0
Trend Line	28	27	1
Moving Average	28	28	0

It can clearly be seen in above table that the value of Gross Enrolment Ratio was accurately predicted by Exponential Smoothing model whereas ARIMA produced the value with a difference of 1.

The values of Transition Rate and Gross Enrolment Ratio fluctuate rarely. The fluctuation only occurs in case of major change in the Education Policy. Therefore, the time series data of Transition rate as well as Gross Enrolment Ratio is free of major fluctuations. Similarly the value of GER increase/decrease slightly due to change in population. Therefore the time series of GER do not have any trend. This case is similar to the case of forecasting the repeaters where Exponential Smoothing method works very well in the absence of variations and trend. Similarly the exponential smoothing method produced close results to actual value in case of Transition Rate and Gross Enrolment Ratio.

The tables from 4.8 to 4.11 present the forecasts results for Enrolment variables and its related indicators. These tables depicts that the ARIMA and Exponential Smoothing techniques performed well except in the case of enrolment where Trend Line method predicted a better results as compared to Exponential Smoothing. But still the value predicted by ARIMA model is closest to the actual value.

In order to better evaluate the Forecasts, the forecasts criteria were changed. The district of Rawalpindi was selected and various disaggregation levels of Enrolment variables were given as input. The output produced by the EDSS is given in the following figure:


Forecast Results

FORECAST RESULTS

Analysis Variable: Enrolment | Province: Punjab | District: RAWALAPINDI | Forecast Model: ARIMA

Base Year: 2005 | End Year: 2011 | Forecast Steps: 1 | Level: Primary

Gender: Both | Location: Rural | Sector: Public | Graph: Actual and Forecasted Values

Forecasted Indicators	2005	2006	2007	2008	2009	2010	2011	2012	Gra
Enrolment	131484	123833	112244	112012	113249	117502	109333	110565	
Population	283100	288000	287000	278000	281000	278000	281000	276000	
OOSC	151616	164167	174756	165988	167751	160498	171667	165435	
Gross Enrolme...	46	42	39	40	40	42	38	40	
Net Enrolment ...	37	34	31	32	32	34	31	32	
Survival Rate	0	69	57	87	96	99	63	90	
Year Input Per ...	0	6.4	7.2	5.6	5.4	5.3	6.5	5.2	
Coefficient of E...	0	78	70	89	93	94	77	96	
Repetition Rat...	0	3	2	2	2	2	2	3	
Repetition Rat...	0	4	2	2	2	2	2	3	
Repetition Rat...	0	4	3	2	2	2	2	3	
Repetition Rat...	0	8	5	4	4	3	2	3	
Repetition Rat...	0	1	0	2	2	2	1	2	
Promotion Rat...	0	86	86	96	100	101	86	86	
Promotion Rat...	0	91	88	97	99	101	88	95	
Promotion Rat...	0	92	87	96	99	98	90	104	
Promotion Rat...	0	78	75	88	90	90	85	94	

Figure 4.12:*Forecast Results of Enrolment Variable*

The values of each indicator from the above figure are presented in the following tables and paragraphs.

The Net Enrolment Rate indicator has the actual value of 31 in the year 2011. The Exponential Smoothing model produced the value of 33 whereas 35 and 34 values were predicted by the ARIMA and Trend Line method respectively. Hence the Exponential Smoothing Model predicted the closes results. The following table shows these comparisons:

Table 4.12: *Forecast results of Net Enrolment Ratio Indicator*

Net Enrolment Ratio			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	31	35	4
Exponential Smoothing	31	33	2
Trend Line	31	34	3
Moving Average	31	33	2

The prediction procedure for Net Enrolment Ratio is similar to the GER as it is derived from the GER. The time series data of NER do not have any trend as well as variations, therefore, Exponential Smoothing is the best candidate for prediction which produced the closes results. Another important indicator in primary education is Survival Rate to Grade 5. This indicator tells us that how many students who entered in class 1 have successfully completed their primary education. The higher the value, the more number of students completed the primary education. In predicting the Survival Rate to Grade 5, the exponential smoothing method produced the closes results with a value of 86. The other two methods produced the poor results.

Table 4.13: *Forecast results of Survival Rate to Grade 5*

Survival Rate to Grade 5			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	63	101	38
Exponential Smoothing	63	86	23
Trend Line	63	110	47
Moving Average	63	98	35

Another indicator related to Survival Rate is Year Input per Graduate. This indicator shows that how many years have been spent by a student to complete the primary education. Ideally the value should be 5 but due to repetition and drop-outs, the value exceeds 5. If the value of this indicator is on high side, then it means that on average a student is spending more years in completing primary education. In predicting the value of this indicator, once again the Exponential Smoothing model produced the most appropriate results. It predicted value of 5.5 against the actual value of 6.5. The following table shows this comparison:

Table 4.14: *Forecast results of Years Input Per Graduate*

Years Input Per Graduate			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	6.5	5.1	1.4
Exponential Smoothing	6.5	5.5	1.0
Trend Line	6.5	5.1	1.4
Moving Average	6.5	5.4	1.1

The Survival Rate and Years Input per Graduate indicators are calculated based on data of enrolment and repeaters for the consecutive two years. This data is further used in Cohort Model to calculate Survival Rate and Years Input per Graduate Indicators. The Cohort Model is very sensitive to repeaters. Therefore, if the repeaters data have fluctuations then these two indicators will also have fluctuations in their predictions. And if the repeaters data is stable then the predictions will also be stable. In the present case, the repeaters data have negligible fluctuations; therefore, Exponential smoothing Method produced the best forecasts for both the Indicators.

The Access to education is very important in assessing that how many children have access to school. Gross Intake Rate is an indicator of access. It tells us that how many children are entering into the first grade of education of a specific level irrespective of their age. The higher value of this indicator shows that more children are entering into school. The following table shows the results of prediction of this indicator which shows that the Exponential Smoothing method produced most appropriate results:

Table 4.15: *Forecast results of Gross Intake Rate Indicator*

Gross Intake Rate			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	42	47	5
Exponential Smoothing	42	44	2
Trend Line	42	44	2
Moving Average	42	42	0

In any level of education, there exist some students which have over-aged according to their grade. For example if the school entrance age of class 1 is set to 5 years, then there will exist some students with the age of 6 or more years due to repetition. In order to get the percentage of students within the range of respective grade, the Indicator of Net Intake Rate is used. The following table shows the results of this indicator:

Table 4.16:*Forecast results of Net Intake Rate Indicator*

Net Intake Rate			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	34	38	4
Exponential Smoothing	34	35	1
Trend Line	34	36	2
Moving Average	34	34	0

In the education system, the role of teacher cannot be denied. In order to measure that how many students are taught by a single teacher, then the Pupil-Teacher Ratio indicator is used. The following table presents the forecast result of this indicator:

Table 4.17:*Forecast results of Pupil-Teacher Ratio Indicator*

Pupil-Teacher Ratio			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
ARIMA	24	23	1
Exponential Smoothing	24	24	0
Trend Line	24	24	0
Moving Average	24	25	1

Since the Pupil-Teacher Ratio indicator has slight increasing/decreasing trend due to decrease in enrolment, therefore, the Trend Line Method can establish good results. The annual growth rate is calculated first in the Trend Line method. This growth rate is assumed to be constant throughout the time series and forecast periods. Since the growth rate calculated by Trend Line method is negligibly small, therefore, no change in the present value is predicted by the Trend Line method. Same results have also been produced by the Exponential Smoothing method on the basis that the previous year value has been given more weightage in forecasting.

After forecasting the variable and indicators for one year, the forecasts were made for two years and more to analyze the results. The following table shows the forecasts result of two years. The observations that were used to perform forecasts were from year 2005 to 2009.

Table 4.18:*Illustration of two years forecast results of Enrolment Variable*

Enrolment					
Model	**Actual Value**		**Predicted Value**		**Mean Absolute Error (MAE)**
	2010	**2011**	**2010**	**2011**	
ARIMA	90,822	93,891	92,202	91,260	**2,006**
Exponential Smoothing	90,822	93,891	88,276	87,031	4,703
Trend Line	90,822	93,891	84,147	82,671	8,948
Moving Average	90,822	93,891	88,100	86,886	4,874

The above table shows that the values for both years were closely predicted by ARIMA model with Mean Absolute Error of 2006 whereas the poorest prediction was made by Trend Line Method with Highest Mean Absolute Error among these three models.

It was found, during the forecasting of Enrolment for a single year, that the ARIMA model produces better results when data has fluctuations. Therefore, when we predict the enrolment for two or more years, the ARIMA model is a strong candidate for forecasting. The results in table 4.18 show that the ARIMA model produced closes results with minimum value of Mean Absolute Error. The reason behind better performance of ARIMA is that it efficiently calculates the variations in the time series and gives weightage of this variation in the future forecasts.

Similar to Enrolment, the Repeaters values for two years were predicted and are presented in the following table

Table 4.19:*Illustration of two years forecast results of Repeaters Indicator*

Repeaters					
Model	**Actual Value**		**Predicted Value**		**Mean Absolute Error (MAE)**
	2010	**2011**	**2010**	**2011**	
ARIMA	1,410	1,235	4,762	4,938	3528

Exponential Smoothing	1,410	1,235	2,818	2,747	1460
Trend Line	1,410	1,235	2,183	2,136	837
Moving Average	1,410	1,235	2,239	2,236	915

This table shows that the closest value of Repeaters was predicted by Trend Line Method as compared to other two methods. The Mean Absolute Error of Trend Line was minimum i.e. 837.

The following table shows the predicted values of Transition Rate for years 2010 and 2011:

Table 4.20:*Illustration of two years forecast results of Transition Rate Indicator*

Transition Rate					
Model	**Actual Value**		**Predicted Value**		**Mean Absolute Error (MAE)**
	2010	**2011**	**2010**	**2011**	
ARIMA	85	88	63	63	24
Exponential Smoothing	85	88	77	82	7
Trend Line	85	88	84	95	4
Moving Average	85	88	79	77	9

Like the Repeaters Indicator, The Transition Rate was best predicted by Trend Line Method with Mean Absolute Error of 4. The Second closer value was predicted by Exponential Smoothing Method which predicted the valueshaving Minimum Absolute Error of 7.

Since the time series data of Transition Rate contains some trend, therefore, Trend Line method produced better results. This method only performs better when the time series data contains increasing/decreasing trend.

Next the indicator of Gross Enrolment Ratio was forecasted. The results as follows:

Table 4.21:*Two years forecast results of Gross Enrolment Ratio indicator*

Gross Enrolment Ratio					
Model	**Actual Value**		**Predicted Value**		**Mean Absolute Error (MAE)**
	2010	**2011**	**2010**	**2011**	
ARIMA	28	29	29	28	1
Exponential Smoothing	28	29	28	27	1
Trend Line	28	29	27	26	2
Moving Average	28	29	33	30	3

It can be seen from the above table that the best value was predicted by ARIMA model with Mean Absolute Error of 1. The next closer value was predicted by Exponential Smoothing model. The exponential smoothing method accurately predicted the value for the year 2010 but the forecasted value for the year 2011 was not accurate therefore it produced the results with same Mean Absolute Error as of ARIMA model.

4.3.2 Comparison of Forecasting Models

In the present study, two forecast models from the Data Mining field have been implemented. These models are Autoregressive Model and Exponential Smoothing. The results obtained from these two models have been compared with the conventionally used Trend Line Model. The results have shown that the performance of the data mining models is far better than the Trend Line Model.

The Autoregressive model has different variations. The simplest model is called Autoregressive Model (AR Model). Another variation, made by addition of Moving Average technique to this method, is called Autoregressive Integrated Moving Average Model (ARIMA Model). If seasonality effect is added in the data, then the Seasonal ARIMA (SARIMA) model will be used.

Autoregressive models attempt to describe the movements in a stationary time series as a function of what are called "autoregressive and moving average" parameters. These are referred to as AR parameters (autoregressive) and MA parameters (moving averages). An AR model with only 1 parameter may be written as:

$$\mathrm{X(t)} = \mathrm{A(1)} * \mathrm{X(t-1)} + \mathrm{E(t)} \qquad \text{Equation … 2.23}$$

Where:

X(t) = time series under investigation

A(1) = the autoregressive parameter of order 1

X(t-1) = the time series lagged 1 period

E(t) = the error term of the model

This simply means that any given value X(t) can be explained by some function of its previous value, X(t-1), plus some unexplainable random error, E(t). If the estimated value of A(1) was .30, then the current value of the series would be related to 30% of its value 1 period ago. Of course, the series could be related to more than just one past value. For example,

$$X(t) = A(1) * X(t-1) + A(2) * X(t-2) + E(t) \quad \text{Equation … 2.24}$$

This indicate that the current value of the series is a combination of the two immediately preceding values, X(t-1) and X(t-2), plus some random error E(t). Our model is now an autoregressive model of order 2.

The Autoregressive model produced promising results in forecasting various variables related to National Education Management Information System. The reason behind this is that this model uses all the available observations for carrying out the forecasting. Hence none of the observations are missed out. Secondly, the autoregressive models workbest when the available data exhibits a stable or consistent pattern over time with a minimum amount of outliers. Since the data of the NEMIS database is already cleaned therefore there are negligible chances of presence of an outlier. Thirdly, the autoregressive models are more suited to stationary time series.Stationary implies that the series remains at a fairly constant level over time. If a trend exists, as in most economic or business applications, then your data is not stationary. Therefore all these three reasons provide the basis for producing a sound forecast.

The Second model implemented in the presented study is called Exponential Smoothing (or Weighted Moving Average). Simple Moving Averages models are used to filter random "white noise" from the data, to make the time series smoother or even to emphasize certain informational components contained in the time series. However, one major drawback of simple moving average method is that it is simply average of previous n terms. Hence if we want to weight a term more than the others, then this method is not suitable. For this purpose the Exponential Smoothing (Weighted Moving Average) Model is used. This exponential smoothing method looks at past data and tries to logically attach importance to certain data over other data. Hence the weighting factor is added to each term. In the case of education system, various policies, plans, projects change over time which results in new results which have more importance therefore this model is best suited for NEMIS time series data. The following lines represent the equation used by Exponential Smoothing method for creating forecasts:

New Forecast = a (most recent observation) + (1 - a) (last forecast)

- OR -

New Forecast = last forecast - a (last forecast error)

where $0 < a < 1$ and generally is small for stability of forecasts (around .1 to .2)

In symbols:

$$F_{t+1}=aD_t+(1-a)F_t \quad \text{Equation … 2.25}$$

$$= a D_t + (1 - a)(a D_{t-1} + (1 - a) F_{t-1}) \quad \text{Equation … 2.26}$$

$$= a D_t + (1 - a)(a)D_{t-1} + (1 - a)^2 (a)D_{t-2} + . . . \quad \text{Equation … 2.27}$$

It is evident from the above equations that the method applies a set of exponentially declining weights to past data. It is easy to show that the sum of the weights is exactly one.

Like the Autoregressive model, this model also uses all the available time series observations. This results in better forecasts as compared to conventional Trend Line model.

The Trend Line method performed very poor in forecasting as compared to these two models. The major reason is that this model uses only two observations from the entire series i.e. first and last observation. All the other observations are ignored. Hence the forecast of the model is dependent on only two observations.

The following points summarize the overall performance of forecast models:

- If the data contains high fluctuations the ARIMA model can be used for forecasting.
- If the time series contains continuous increasing/decreasing trends, the Trend Line method is best option for forecasting.
- The Exponential Smoothing Method can be used when the future values are likely to depend on the most recent values.
- The Exponential smoothing method can also be used for the time series containing fluctuations, but in this case the moving average order should be increased. However, this cannot be applied if we have the lesser number of observations.
- The Trend Line method performs poor if the time series have fluctuations because this method does not has the capability to remove fluctuations in the time series.
- The ARIMA model can also produce good results if time series have increasing/decreasing trend. Since increasing/decreasing trend is a special case of fluctuation, therefore, the ARIMA model has the ability to cater these fluctuations.
- If lesser number of observations are available, then all the observations have much weightage because neglecting the each value can cause unwanted results. Therefore in this situation the Exponential Smoothing method produce better results.
- The Exponential Smoothing method is also best used for short-term forecasts.

4.3.3 Comparison Results

Since the EDSS tool also provide comparisons of various districts, therefore, it is pertinent to discuss the results of these comparisons. Following figure shows the comparison results.

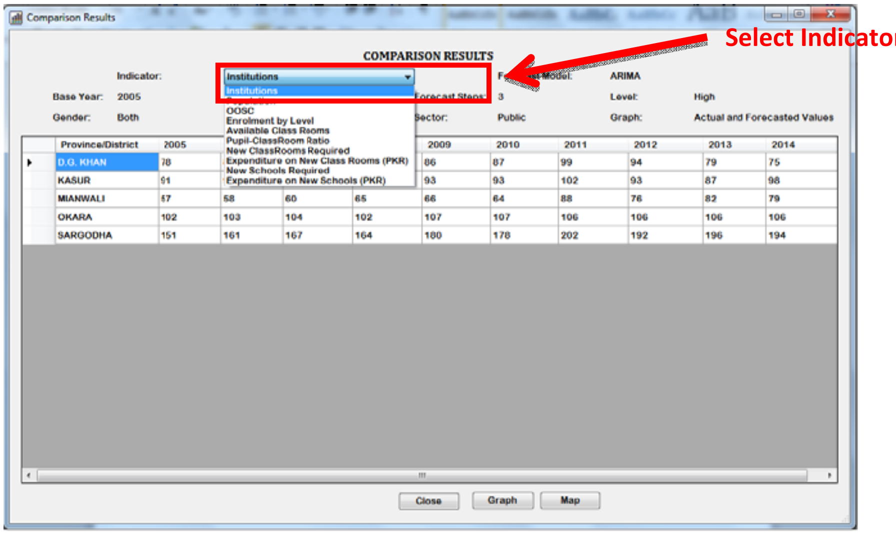


Figure 4.13:*Comparison Results of Institutions variable for 5 districts*

The user needs to select the variable or its related indicators from the drop down list available in the upper portion of the output screen. The actual and predicted values of selected variable/indicator will be shows in the lower portion of the screen.

As charts and graphs are very important to understand the results graphically, therefore, the EDSS tool has the capability to provide the comparison charts. If the user wants to view the graph showing the comparison results, then it can simply be produced by clicking the Graph button. The following figure shows the graph generated by EDSS after performing comparison.

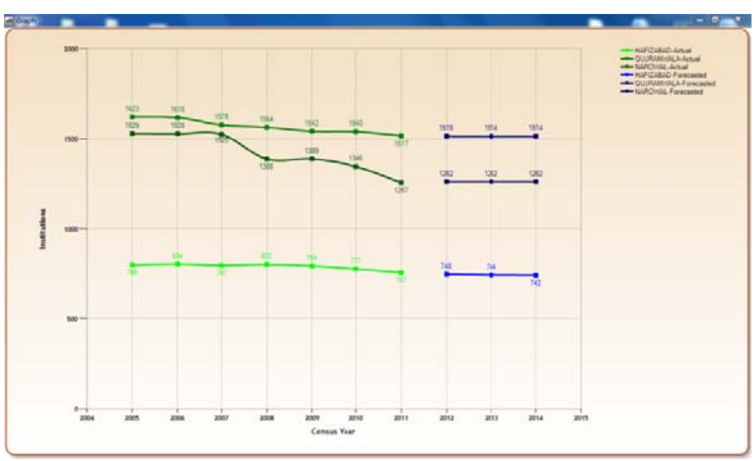


Figure 4.14:*Comparison Results Graph of Institutions variable for various districts*

The values in the shades of green color represent the actual values whereas the values shown in the shades of blue color are predicted values. These graphs are widely used to analyze the trends in education as well as for comparison of various districts and provinces. The Census years are shown along horizontal axis whereas the vertical axis represents the values of selected variable/indicator.

4.3.4 Top-N Analysis Results

The EDSS tool also has the functionality of providing top/bottom –N analysis of each indicator and variable. The education manager can analyze any number of top and bottom districts/provinces for a particular year and indicator.

The following figure depicts the Top-N Analysis result of institutions variable for the year 2014.

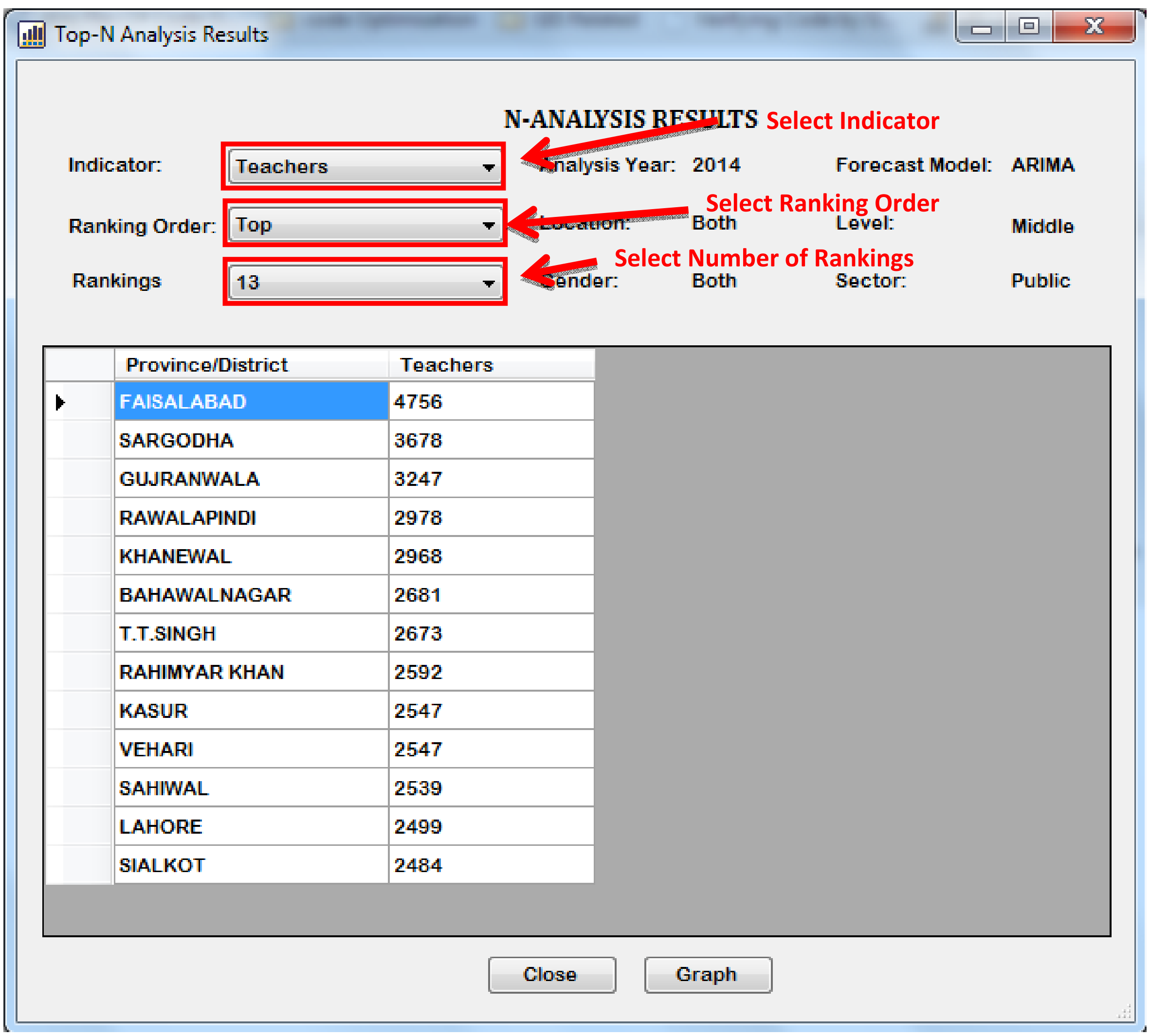


Figure 4.15:*Top-N Analysis Results*

The top/bottom-N analysis is normally used in situations where the policy or decision makers intend to analyze the best or worst perform district/provinces. The top-N analysis in EDSS is customized in order to cater the requirements of policy makers. First of all the user needs to select variable/indicator for which the analysis is required. Secondly, the ranking order, i.e. Top or Bottom for best or worst performing Districts/Provinces respectively, is selected. In the third and last step, the number top/bottom rankings are selected.

In order to visually understand the Top-N analysis, these results can be viewed graphically with the help of chart. Following is the graph generated by EDSS tool on the basis of Top-N Analysis:

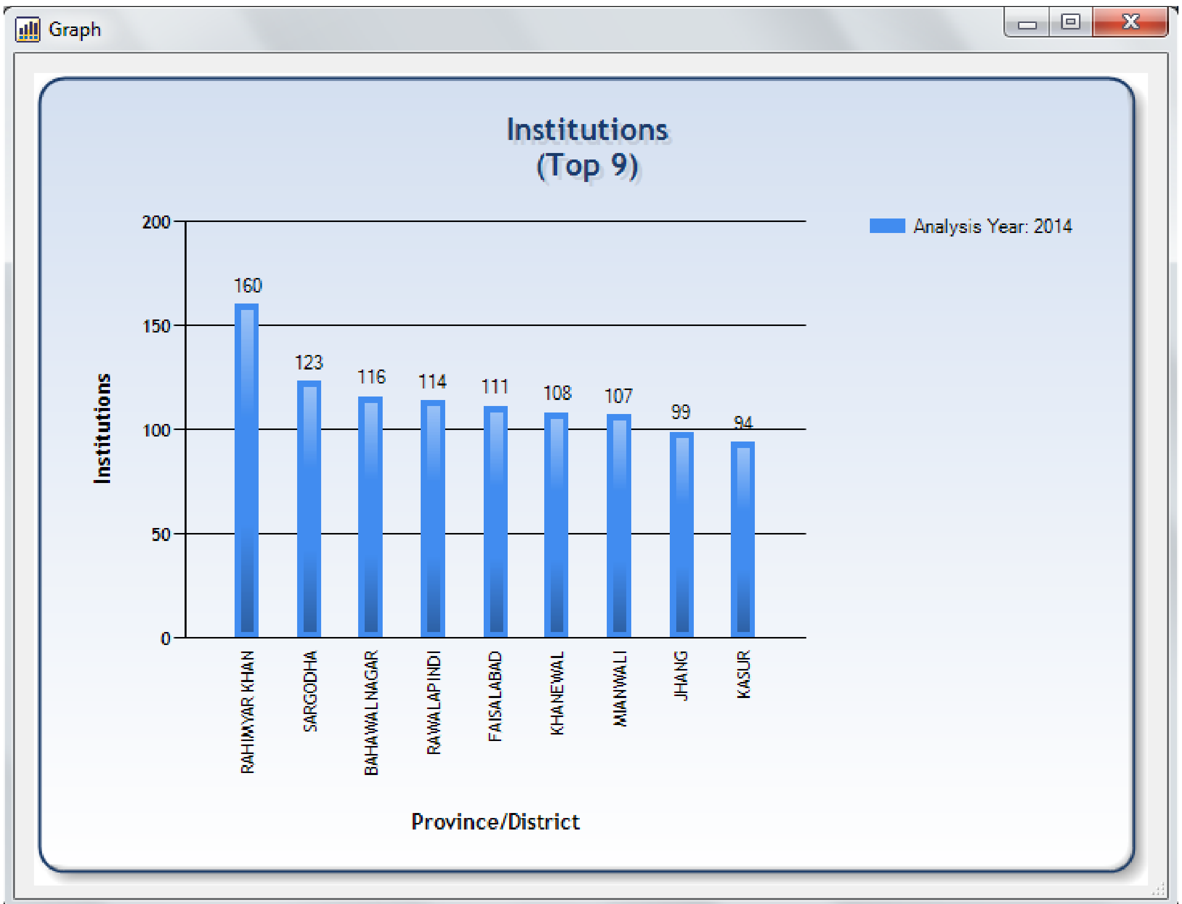


Figure 4.16:*Top-N Analysis Graph*

The Provinces and Districts are represented along horizontal axis whereas the values of variable or indicator are projected along the vertical axis. The above chart shows that the district Rahim Yar Khan will be having most number of schools in rural areas in 2014.

4.4 Multivariate Forecasting

Usually, multivariate forecasting methods rely on models in the statistical sense of the word. However, one may classify multivariate methods with regard to whether they are a theoretical, such as time-series models, or structural or theory-based.

4.4.1 Is multivariate better than univariate?

Multivariate methods are very important in economics and much less so in other applications of forecasting. In standard textbooks on time-series analysis, multivariate extensions are given a marginal position only. Empirical examples outside economics are rare. Exceptions are data sets with a predator-prey background, such as the notorious data on the population of the Canadian lynx and the snowshoe hare. In contrast, the multivariate view is central in economics, where single variables are traditionally viewed in the context of relationships to other variables. Contrary to other disciplines, economists may even reject the idea of univariate time-series modeling on grounds of the theoretical interdependence, which appears to be an exaggerated position.

In forecasting, multivariate models are not necessarily better than univariate ones. While multivariate models are convenient in modeling interesting interdependencies and achieve a better (not worse) fit within a given sample, it is often found that univariate methods outperform multivariate methods out of sample. Among others, one may name as possible reasons:

1. Multivariate models have more parameters than univariate ones. Every additional parameter is an unknown quantity and has to be estimated. This estimation brings in an additional source of error due to sampling variation.
2. The number of potential candidates for multivariate models exceeds its univariate counterpart. Model selection is therefore more complex and lengthier and more susceptible to errors, which then affect prediction.
3. It is difficult to generalize nonlinear procedures to the multivariate case. Generally, multivariate models must have a simpler structure than univariate ones, to overcome the additional complexity that is imposed by being multivariate. For example, while a researcher may use a nonlinear model for univariate data, she may refrain from using

the multivariate counterpart or such a generalization may not have been developed. Then, multivariate models will miss the nonlinearities that are handled properly by the univariate models.

4. Outliers can have a more serious effect on multivariate than one univariate forecasts. Moreover, it is easier to spot and control outliers in the univariate context.

An additional complication is conditional forecasting, which means that a variable Y is predicted, while values for a different variable X are assumed over the prediction interval. If X is a policy variable, it may make sense to regard it as fixed and not to forecast it. Even then, true-life X may react to future values of Y, while that reaction is ignored in the conditional forecast. The arguments of 'feedback' by Chatfield and the so-called Lucas critique are closely related. If the forecasting model is designed such that forecasts of one or more variables of type X are not generated, while one or more variables of type Y are modeled as being dependent on X, the system is called open-loop. If all variables are modeled as dependent on each other and on lags, the system is called closed-loop. Even closed-loop systems may allow for deterministic terms, such as constants, trend, or other variables that can be assumed as known without error at any future time point.

4.4.2 Regression Analysis

Regression analysis is widely used for prediction and forecasting. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships

A large body of techniques for carrying out regression analysis has been developed. Familiar methods such as linear regression and ordinary least squares regression are parametric, in that the regression function is defined in terms of a finite number of unknown parameters that are estimated from the data. Nonparametric regression refers to techniques that allow the regression function to lie in a specified set of functions, which may be infinite-dimensional.

In the past chapters the univariate forecasting has been presented. Now, the multivariate forecasting is discussed in the coming paragraph. The Regression analysis method is used as primary method for knowing the relationship among the variables.

For the experiment, four variables were taken as inputs that are Institutions, enrolment by level, Out of school children and Pupil-Classroom Ratio. The enrolment by Level was taken as Independent variable as the value of this variable has to be predicted. The other three variables were taken as dependent variables.

At the next step, the relation of independent variable with dependent variables was analyzed and the values of coefficients were calculated. On the basis of these coefficients, the equation was made to predict the future values. Following are the results of calculating the coefficients using Regression analysis:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.99549189
R Square	0.991004103
Adjusted R Square	0.977510258
Standard Error	1236.543717
Observations	6

	Coefficients
Intercept	-35.88425767
Institutions	150.6464872
OOSC	0.029959596
PCR	-6.911433557

RESIDUAL OUTPUT

Observation	*Predicted Enrolment by Level*
1	94570.78716
2	93705.45591
3	87086.96429
4	83433.99888
5	76320.95065
6	75647.84311

On the basis of these coefficients the equation was developed that will be used for forecasting. Following is the equation:

Enrolment by Level=-35.88 + 150.646*Institutions+0.0299*OOSC-6.911*PCR

Using this equation, the value of Enrolment for the year 2011 was predicted. Following are the result of this forecast:

Table 4.22:*Illustration of forecast results of Enrolment variable using Regression Analysis*

Enrolment by Level			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
Regression Analysis	74635	76,923	2288

The actual values of Institutions, Out of school children and Pupil-Classroom Ratio were used to predict the value of Enrolment by Level for the year 2011.

If we compare this result with the results of univariate time series analysis and forecasting methods presented in Table 4.3, it is evident that the univariate methods produced much better results. The multivariate time series analysis produced results nearer to trend-line method.

Similarly the forecasts for two years were made. Following table shows the forecast results:

Table 4.23:*Two years Forecast Results of Enrolment by Level using Regression Analysis*

Enrolment by Level					
Model	**Actual Value**		**Predicted Value**		**Mean Absolute Error (MAE)**
	2010	**2011**	**2010**	**2011**	
Regression Analysis	74,917	74,635	75,648	76,923	1510

When the results of above table was compared with univariate forecasting models (table 4.7), it was found that this method produce good results. But still the best forecast was made by Exponential Smoothing.

Another experiment was held to made forecast of repeaters on the basis of Enrolment, Transition Rate, and Gross Enrolment Rate. The data for the years 2005 to 2010 was used as baseline to construct the Regression Analysis Equation. Following are the coefficient values:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.165731947
R Square	0.027467078
Adjusted R Square	-1.431332304

Standard Error	2855.718951
Observations	6

	Coefficients
Intercept	13477.2112
Enrolment	-0.100814655
Transition Rate	-7.05825375
Gross Enrolment Ratio	-25.17670956

RESIDUAL OUTPUT

Observation	*Predicted Repeaters*
1	3372.513541
2	3415.238326
3	2972.934965
4	2833.288806
5	3605.901186
6	3016.123176

On the basis of above table, the equation was developed. Following is the equation:

Repeaters = 13477.211-Enrolment*0.101-Transition Rate*7.058-GER *25.177

Following are the results produced by this equation:

Table 4.24:*Illustration of forecast results of Repeaters Variable*

Repeaters			
Model	**Actual Value**	**Predicted Value**	**Difference (in absolute numbers)**
Regression Analysis	2,448	2,686	238

If we compare this result with the table 4.9 then it is found that still the univariate model produced the better results.

It is pertinent to mention here that the forecast of Enrolment and Repeaters variables using regression analysis was made on the basis of three variables. In the above tables, the actual values of these three variables were used in forecasting. So if the predicted values are used instead of actual values, then the chances of errors will be higher. Therefore, in EMIS the univariate forecasting models are normally applied.

Chapter 5

DISCUSSIONS

The most striking feature of Pakistan's education system is its inherent inequalities. With a net primary enrolment rate of 68%, there are an estimated 7 million primary-aged children are out-of-school, mostly due to poverty and partly due to non-existence of school (as some live in remote or far flung areas), absence of teacher and/or learning material or other reasons. Of the children who do attend school, given that most (63%) of the population lives in rural areas in relatively modest conditions, do not obtain quality education. This is mainly due to the co-existence of parallel streams of primary and secondary schooling, further divided across publicand private arrangements, catering to different socioeconomic classes in the country. The average percentage of population that ever attended school is higher for urban (73%) than rural (50%) areas. There are also wide variations across genderwithin urban as well as rural areas with males enjoying a higher access than females.

In Pakistan, recent statistics for the public sector institutions show that there are 163,892 formal schools offering education till the higher secondary level i.e., grades XI and XII. Of these schools, 48% are boys' schools, 34% are girls' schools and 18% are mixed schools.

Public formal school enrolment is approximately 21.5 million students, with 50%-60% males at every level of education. Of these, over one-half are primary level students, followed by pre-primary students, who constitute around 20% of all students. Contribution of the private sector to education is estimated at about one-third of total school enrolments.

Progress in education is best reflected by levels of, and increase in, enrolment rates. In Pakistan, enrolment rates at all education levels are low and annual changes are marginal. For example,over the six year period i.e. 2004/05 to 2010/11, gross primary enrolment and net primary enrolment rates for age group 5-9 years, both progressed by only 1% point per annum. Gross enrolments rose from 86% to 92%, while net enrolments rose from 52% to 68% in 2011/12.

Pakistan is likely to lag behind in attaining universal primary education by 2015 if comprehensive efforts are not undertaken urgently. The preceding situation analysis of its education sector reveals that on average, for the country as a whole, 32% of the primary age (5-9 year olds) are out of school, with wide variance across provinces, sex and urban-rural location.

School enrolment does not guarantee completion of primary schooling. For Pakistan, estimated information reveals that of all the children entering primary school, 70% reach Grade 5, though with a narrow difference between boys (71%) and girls (68%). The projections, based on past trend, suggest that completion/survival rates will not exceed 80% by 2015/16. Efforts need to be invested to enhance completion/survival rates to achieve UPE.

For primary age children, statistics yield a total number of 6.7[1] million children out of school. Access to education is a major challenge in Pakistan which is influenced by in-come distribution, urban-rural residence as well as male-female differences. The average percentage of population that ever attended school is higher for urban (73%) than rural (50%) areas. There are also wide variations across gender within urban as well as rural areas with males enjoying a higher access than females.

Out of school children can be broadly classified into two groups: (i) children who have never attended school; and (ii) children who dropped out of school before completing primary education.To improve access to, and retain children in, education, these factors need to be addressed forcefully. In every province, children in poor rural neighborhoods in most deprived districts will have to be targeted.

In 2011/12, Pakistan allocated Rs. 390 billion (2% of its GDP) on education. By international standards, this is a low proportion, though, given the limited absorptive capacity of the education sector, only 90% of it was eventually utilized. The pre and primary sub-sector spent about Rs. 138 billion (39% of total education expenditure) which made it the single most important sub-sector. This was followed by Rs 119,642 million by secondary (34%) and Rs. 74,935 million by tertiary (21%) sub-sectors.

At the national level, primary education expenditure per primary student is estimated at Rs 14,954, almost over twice of education expenditure per child, aged 5-9 years, which is calculated as Rs 6,500.

Preliminary rough calculations yield a total amount of Rs 313.5 billion for enrolling 7 million out of school children over the next three years. This estimate is based on the assumption of current formal public school expenditure per primary school student. As new entrants would

[1] According to Pakistan Social and Living Standards Measurement Survey (PSLMS), there are 9.3 million out of school children in Pakistan

be distributed across formal and non-formal schools, these costs will change, perhaps likely to decline as non-formal instruction is somewhat less expensive than formal schooling.

5.1 Primary Enrolment Rates

On average, Pakistan's gross primary enrolment rate (GER) is 85%, with 92% for boys and 79% for girls. KP displays the highest GER of 101%, followed by GB (95%), Punjab (88%), ICT (87%) and Sindh (79%).

In Pakistan, of all the primary-aged (5-9 years) children, 68% are enrolled in primary school. However, wide variance is displayed across province, sex and location.The highest net primary enrolment rate is in KP (81%) where 92% of all boys (aged 5-9 years) and 68% of all girls (aged 5-9 years) attend primary school. This is followed by GB (76%); Punjab (70%); and ICT (70%).

About two-thirds of children attend school in Sindh (63%) and FATA (62%) while only one-half children in Baluchistan (51%) are enrolled in primary schools. ICT is also the only area in the country where primary enrolment rate is higher for girls (72%) than boys (68%) while AJ&K has almost an equal enrolment rate (58%-59%) for boys and girls. Also, there are some discrepancies in net primary enrolment rates across urban-rural location, with urban areas showing higher net primary enrolment than rural areas. While the urban-rural differences are small in three provinces and GB, but these are more pronounced in Baluchistan, ICT and AJ&K. One interesting finding is that net enrolment rates are higher in rural Sindh (65%) than in urban Sindh (61%).

5.2 Survival to Grade 5

School enrolment does not guarantee completion of primary schooling. Therefore, one needs to assess the survival rates and/or completion rates of primary school children. For Pakistan, estimated information reveals that of all the children entering primary school, 70% reach Grade 5. For boys this rate (71%) is slightly above than that for girls (68%). Among the provinces and areas, the highest rate of survival is for ICT (91%) while GB (32%) is lowest. In Sindh, FATA and KP, almost two-thirds of the children reach grade 5 while in Baluchistan only one-half survive up to the final primary class.

There are differences between urban and rural survival rates. For Pakistan urban survival rate (72%) is higher than the rural rate (69%). For most provinces and areas, urban survival rates are higher than rural rates except in Baluchistan, Punjab and AJ&K where rural survival rates are higher than urban rates.

Of the children who do attend school, given that most (63%) of the population lives in rural areas in relatively modest conditions, do not obtain quality education. This is mainly due to the co-existence of parallel streams of primary and secondary schooling, further divided across public and private arrangements, catering to different socioeconomic classes in the country.

5.3 Out of School Children

Out of school children can be broadly classified into two groups: (i) children who have never attended school; and (ii) children who dropped out of school before completing primary school.

Out of school children in Pakistan comprised of two different categories i.e. left outs (those who never enrolled) and dropouts (those who enrolled but dropped out before completion of primary education due to various reasons). The main source of educational data and statistics in Pakistan is National Education Management Information System (NEMIS)

NEMIS gives a complete picture of net enrolment and out of school children for all provinces and federating units for both the public sector (formal and non-formal) and private sector.

The total number of out of school children at primary level as per NEMIS statistics is 6.2 million in all the four provinces and 6.7 million all over Pakistan including AJK, ICT, GB and FATA. The said number has been calculated by subtracting net enrolment of 14.5 million out of primary age group ($5\text{-}9^{+}$) population of 21.2 million. The province wise total number of out of school children in all the four provinces is 3.2 million in Punjab, 1.8 million in Sindh, 0.55 million in KPK and 0.54 million in Baluchistan.

Out of school boys: The overall population of primary age group boys in Pakistan is 11 million. Out of these 8.1 million are enrolled. Whereas remaining 2.9 million are out of school. Half of the out of school children live in Punjab province.

Out of school girls: The overall primary age group population of girls is 10 million. Out of these 6.4 million are in school and remaining 3.78 million are out of school. As in case of boys almost half of the out of school girls live in Punjab province.

5.4 Private Sector Contribution in Educational Development

Private Sector is playing an important role in the promotion of education in Pakistan. NEMIS data indicates that presently, (in 2011-12) there are 18,208 private primary schools; 25,788 middle/lower secondary schools; and 17,388 high schools. In percentage terms, 12% primary; 61% middle/lower secondary and 61% high schools are in private sector respectively.

At primary level overall 4.8 million (34%) children of $5\text{-}9^{+}$ age group against around 14.4 million total net enrolment are enrolled in private sector schools. 2.75 million (34%) boys against total 8.06 million are in private schools whereas 2.1 million girls i.e. (33% of total) are studying in private schools. Private sector enrolment is increasing because of overall better quality of education as compared to public sector. Province wise the highest number of children i.e. 42% of the total are in Punjab studying in private sector followed by AJK 36%, Sindh 26% and KP and Gilgit-Baltistan 22% each. The lowest private sector contribution in terms of primary education net enrolment is in Baluchistan i.e. 13%, FATA 18%, and ICT 19%.

5.5 Transition Rate

The Transition Rate indicator is used to assess the number of students moving from grade to another. Ideally this value should be 100 but due to repetitions and drop outs this value remains less than 100. It is found that in every district the transition rate in initial grades of a level of education is high which gradually decreases due to repetitions and drop-outs. The transition rate in urban areas is better as compared to rural areas. Furthermore, the transitions rates of private sector are far better than the public sector institutions.

5.6 Pupil Classroom Ratio

This indicator shows that how many students are accommodated in a single classroom. In Pakistan, hundreds of schools are shelter-less where students are getting education in open-air. This results in high value of Pupil Classroom Ratio. Moreover, in many districts of the country, the school administration runs multiple classes in a single classroom due to non-availability of classrooms. This results in high value of Pupil-Classroom Ratio.

5.7 Pupil Teacher Ratio

The Pupil-Teacher Ratio specified that to how many children a teacher is teaching. If the value of this indicator is high, it means that a single teacher is teaching many students. In Pakistan, the standard value of PTR is 30. In urban areas of country where enrolment is very high, this value exceeds 40 in some districts. Furthermore, increase in student enrolment and non-availability of teachers is also a major reason in increased value of this indicator. Furthermore, the private sector employ less teachers due to financial constraints, therefore, the value of PTR in private sector is very high as compared to public sector. The value of PTR is high at primary level but it decreases at middle and high level due to the reason that students do not enroll themselves in middle or high school education.

5.8 Enrolment

Enrolment refers to the number of students admitted in a school. In Pakistan, the public sector institutions are providing education throughout the country irrespective of locality. Due to the evolution of private sector and better facilities provided by this sector, the trend has been set to move the students from public to private sector. Due to this the enrolment in public sector is decreasing whereas in private sector, it is increasing. Furthermore, the private sector has better setup in urban areas whereas the rural areas are neglected. However, the enrolment at middle and high levels are increasing in public sector.

5.9 Repetition Rates

The study has found that the Repetition Rates are increasing annually. A few years back the repetition rates were nearly 1 to 2 percent. Due to poor teaching standards, it increased upto 6 percent in rural areas of some districts. This shows the weak internal efficiency of the education system especially at primary level.Some serious attentions are required in this area in order to improve the situation.

5.10 Years input per Graduate

This indicator is another measure of internal efficiency of the system. This indicator tells that how many years have been spent by the student to complete the primary level of education. Since the Survival Rate is declining and drop-out rate as well as repetition rate is increasing therefore the value of this indicator is also increasing. The value of this variable is more increased in rural areas as compared to urban areas.

5.11 Share of Education in Total Expenditures

As percentage of total expenditures at the national level, actual education expenditures have remained more or less constant for the past three years, remaining within a narrow band of 7.7% to 8.1%. For the provinces, this percentage i.e. education expenditures as percentage of total provincial expenditures, is much higher i.e. roughly around 15%-20%. Given huge defense, energy and interest expenditures, the total federal expenditures are very high and though federal education expenditures are substantial but as a percentage of total expenditures these appear small i.e. ranging between 1.7% and 2.3%.

5.12 Education Policy and Plans in Pakistan

During the past decade, there were two major education policy interventions- the National Education Policy 1998-2010; and the National Education Policy 2009. These policies were developed with the consensus of a wide group of stakeholders i.e., federal and provincial government educational officials, education experts, academia, private organizations, Non-Government organizations and international development partners.

5.12.1 The National Education Policy (1998-2010)

The National Education Policy (1998-2010) emphasized increased enrolments in public sector schools and higher budgetary allocations to education. It advocated the removal of urban- rural and gender imbalances, improving the quality of education at all levels particularly through curriculum reform, strengthening education facilities, encouraging private sector participation and effective community involvement. It specifically envisaged addressing the issue of out of school children and expansion of non formal education to complement the formal system; and the implementation of literacy and functional literacy programs for adults.

In light of the National Education Policy 1998-2010, the Education Sector Reforms (ESR: 2001-06), focusing on literacy, universal primary education of good quality, enhanced budgetary allocations (to 3% of GNP), improved technical and higher education, and greater public-private partnership, was developed with all the principal actors of EFA and other stakeholders. The consultation process lasted for more than six months. The organizations which participated included the education departments of all provincial and federating units, non-governments organizations, private sector, UN agencies and bilateral and multilateral donor agencies. In this context, a National Conference on Education for All was held with the objective to launch EFA movement, share EFA goals and strategies with stakeholders and prepare a framework for the national and provincial plans of action. This was followed by a series of discussions and workshops to develop action plans, each based on active participation by all stakeholders i.e., government officials, education experts, NGOs, private sector, and international development partners.

The EFA National Plan of Action (2001-15) was prepared and endorsed by the first Poverty Reduction Strategy Paper (PRSP I 2003-06) but could not be implemented due to lack of financial support, both domestic and external. Similarly 15-year provincial and district EFA plan were prepared.

5.12.2 The National Education Policy (2009)

The current National Education Policy (NEP 2009), developed after several rounds of deliberations with relevant stakeholders, addresses all the dimensions of Pakistan's education sector. The NEP (2009) recognizes that there are close links between equity in educational opportunities and equitable income distribution and income growth. If the education system is constructed on a divisive basis, the divisions it creates can endanger long run economic growth as well as stability of society.

Affirmation of commitment to Pakistan's egalitarian education vision in the service of all citizens and as a driver of economic and social development can help produce a virtuous circle of high level of human and social capital leading to equitable economic growth and social advancement.

The NEP 2009 document identifies policy actions in pursuit of two overriding objectives: (i) widening access to education; and (ii) improving quality. Key policy actions identified are as follows:

1. Provinces and Area Governments shall affirm the goal of achieving universal and free primary education by 2015 and up to class 10 by 2025.
2. Provincial and Area Governments shall develop plans for achieving these targets, including intermediate enrolment targets and estimates of the required financial, technical, human and organizational resources.
3. The plans shall also promote equity in education with the aim of eliminating social exclusion and promoting national cohesion. Greater opportunities shall be provided to marginalized groups of society, particularly girls.
4. To achieve the commitments of Government of Pakistan towards Education for All (EFA) and the MDGs, inclusive and child-friendly education shall be promoted.
5. Special measures shall be adopted to ensure inclusion of special persons in mainstream education as well as in literacy and Technical and Vocational Education (TVE) programs.

6. Governments shall improve quality of educational provision at all levels of education.
7. National Standards for educational inputs, processes and outputs shall be determined. A National Authority for Standards of Education shall be established. The standards shall not debar a provincial and area government/organization from having its own standards higher than the prescribed minimum.
8. Provincial and district governments shall establish monitoring and inspection systems to ensure quality education service delivery in all institutions.
9. Steps shall be taken to make educational provision relevant for the employment market and for promoting innovation in the economy.
10. Universities and research institutes shall place greater emphasis on mobilizing research for promoting innovation in the economy.
11. Educational inputs need to be designed with a comprehension of the challenges and opportunities related to globalization. Strategies shall be developed to optimize opportunities and minimize the potentially negative impacts.

5.12.3 Provincial Education Sector Plans

The provinces of Baluchistan and Khyber Pakhtunkhwa have recently prepared Education Sector Plans. The KP Education Sector Plan (2010-15) aims to achieve Universal Primary Education (UPE) by 2015 ensuring that all boys and girls complete full course of primary education; promote gender equality, achieve quality basic education for all (EFA goal); achieve 50% improvement in the levels of adult literacy, especially for women (EFA goal); introduce government-financed private school subsidizing for areas with low female enrolments; and reduce rural and urban disparities in education. The Baluchistan Education Sector Plan (2013-14 to 2017-18) focuses on: (i) improving quality of education; (ii) early childhood education; (iii) access & equity; (iv) governance & management; and (v) adult literacy & non formal education.

To increase enrolment and retain students, several incentives have been extended for over a decade which include (i) free textbook being provided by the government to the students studying in public sector schools (both formal and non-formal). Students of non-formal schools are given notebooks and writing material in addition to the textbooks; (ii) no tuition fee is charged from the primary education students studying in government schools. (iii) Female students of middle and high school in rural areas of few selected districts are being given monthly scholarships/stipends. Also, free edible oil for high attendance level is being given to both girl students and their teachers; and (iv) provision of free lunch under *Tawana Pakistan* was also initiated in selected schools but had to be discontinued due to (mis)management issues.

5.12.4 Education Sector Reforms Programs

During the decade of 2000, serious efforts were made in three provinces i.e., Punjab, Sindh and KP to implement their respective Education Sector Reforms Programs. The key features of these programs was to provide missing facilities (such as shelter, boundary wall, toilets, etc) in primary schools; provision of free textbooks, stipends, textbook reforms, merit based teacher recruitment, etc. Besides, a number of development programs and projects were launched by the government in collaboration with international development partners to expand the access and improve the quality of education. Under these development projects number of new primary schools have been opened or upgraded to middle (elementary) level.

5.13 Impact of Household Incomes on School Enrolment

Income distribution, urban-rural residence, as well as male-female differences determine the varying access to education (Table 5.1). The average percentage of population that ever attended school is higher for urban (73%) than rural (50%) areas. There are also wide variations across gender within urban as well as rural areas with males enjoying a higher access than females. A review of access to education by income quintiles (the first quintile representing lowest income levels and fifth quintile reflecting the highest income levels) reveals, for all areas and both genders, a positive correlation between income and percentage of population ever attended school.

Table 5.1:*%of Population Ever Attended School by Income Class, Location and Gender*

	Urban			**Rural**		
Income Bracket	**Male**	**Female**	**Both**	**Male**	**Female**	**Both**
PAKISTAN	81	65	73	65	36	50
1st Quintile	56	40	48	48	20	34
2nd Quintile	67	50	59	60	26	43
3rd Quintile	75	58	67	67	38	52
4th Quintile	84	67	75	74	46	60
5th Quintile	93	81	87	83	58	71

Source: PSLMS 2007-08

Public schools play an important role in the low-income rural sections of the society where

85%-90% of children attend public schools while among the upper classes in both urban and rural areas only 21% to 35% attend public schools (Table 5.2). This suggests that investment in the improvement (in school environment, facilities, teaching quality, provision of missing facilities and teaching-learning materials, etc) of public schools can considerably enhance school enrolments, retentions and education quality indicators.

Table 5.2:*Economic Status of Parents by GER (5-9) in Public and Private Sector*

	Urban		Rural	
	Public	**Private**	**Public**	**Private**
PAKISTAN	43	55	74	26
1st Quintile	74	24	90	8
2nd Quintile	60	39	85	14
3rd Quintile	51	48	71	27
4th Quintile	35	63	64	35
5th Quintile	21	77	35	64

Source PSLMS 2007/08

5.13.1 Out of School Children

Among children who never attended school, the most predominant reason for girls is "parents did not allow" (40%), followed by "too expensive" (16%), "child not willing (10%) and "too far" (9%). For boys, the reasons seem a bit different, with "child not willing" (37%), followed by "too expensive" (21%); and "has to help at work" (10%).

Table 5.3:*Reasons for Never Attending School (10-18 years)*

	Boys			Girls		
	Urban	**Rural**	**Overall**	**Urban**	**Rural**	**Overall**
Parents didn't allow	4	6	6	36	41	40
Too expensive	26	19	21	29	13	16
Too far	2	7	6	3	10	9
Education not useful	2	1	1	1	1	1
Had to help at work	9	11	10	2	4	4
Had to help at home	3	3	3	8	7	7
Child not willing	37	36	37	11	10	10
Other	17	16	17	11	14	13

Source: PSLMS 2007-08

For girls, the predominant reason for dropping out of primary school is “child not willing” (14%), followed by “parents did not allow” (10%) and “too expensive” (7%). With boys, too, the predominant reason is also “child not willing” (26%) but this is followed by “had to help at work” (7%).

Table 5.4:*Reasons for Leaving School before completing primary education (10-18 years)*

	Boys			**Girls**		
	Urban	**Rural**	**Overall**	**Urban**	**Rural**	**Overall**
OVERALL	15	37	53	12	35	47
Parents didn’t allow	0	2	2	3	8	10
Too expensive	2	2	4	3	4	7
Too far	0	0	0	1	2	3
Education not useful	0	0	0	0	0	0
Had to help at work	2	7	8	0	1	1
Had to help at home	0	1	1	1	2	4
Completed desired education	0	1	1	0	0	0
Child not willing	8	19	26	3	11	14
Other	3	6	9	2	7	9

Children aged 10 - 18 years that cited the reason indicated for leaving school expressed as a percentage of all children aged 10 -18 years that left school before completing primary level.
Reasons for leaving school before completing primary level: “Other” includes ‘Poor teaching/behaviour’, ‘No female staff’, ‘No male staff’, ‘Child sick/handicapped’, ‘Child too young’, ‘Lack of documents’, ‘Marriage', ‘Service’, and ‘Other’.
Source: PSLMS 2007-08

5.13.2 Education Finance

Pakistan’s education expenditure as percentage of GDP has varied between 1.7% and 2.5% (Table 5.5). Even with these small amounts, the utilization rates have remained at an average of 90%.

Table 5.5:*Education Expenditure as % of GDP*

2003/04	**2005/06**	**2006/07**	**2007/08**	**2008/09**	**2009/10**	**2010/11**	**2011/12**
2.2	2.4	2.42	2.49	2.1	2.05	1.8	2.0*

**Estimated*
Source: Economic Survey of Pakistan (various issues)

5.13.3 Total Education Budgetary Allocations and Expenditures

In 2011/12, over Rs 390 billion were allocated to education and over Rs 350 billion were spent in the sector, yielding a utilization rate of about 90%.

Reviewing the past three years, statistics show that Pakistan’s total budgetary allocations to education, in nominal terms, have been rising and so have the actual total education expenditures (Table 5.6). However, with education's devolution to the provinces under the 18th Amendment of the Constitution during April 2010, federal allocations gradually declined in 2010/2011 and 2011/12 though expenditures increased in 2010/11 before slipping back to the 2009/10 levels. Most provincial allocations and expenditures have been increasing substantially.

Table 5.6:*Education Total Budgetary Allocations and Expenditures (Rs m)*

	2009/10		2010/11		2011/12	
	Budget	**Actual***	**Budget**	**Actual**	**Budget**	**Actual**
Punjab	127804.5	106033.1	164462.3	140161.6	183042.8	166829.9
Sindh**	71949.0	52870.3	78704.1	72394.7	73886.9	47213.0
KP	39495.0	39138.3	47720.4	51030.4	58248.2	67381.2
Baluchistan	4055.8	7380.1	18986.6	19408.5	21745.9	23981.1
Federal	62536.4	48481.5	55665.7	60487.8	53512.5	48150.7
NATIONAL	305840.8	253903.5	365539.1	343483.0	390436.3	353555.9

**Actual: Total actual education expenditures at provincial and district levels.*
*** Estimated*
Source: Office of the Controller General, Accounts (CGA), 2012, Govt. of Pakistan

Overall utilization rates i.e. total education expenditure as % of budgetary allocations at the national level varied between 83% and 94%. Utilization rates above 100% may be due to high increases in staff salaries or additional/supplementary grants from other sources (e.g. Federal Government, donors etc.).

5.13.4 Share of Education in Total Expenditures

As percentage of total expenditures at the national level, actual education expenditures have remained more or less constant for the past three years, remaining within a narrow band of 7.7% to 8.1% (Table 5.7). For the provinces, this percentage i.e. education expenditures as percentage of total provincial expenditures, is much higher i.e. roughly around 15%-20%. Given huge defense, energy and interest expenditures, the total federal expenditures are very high and though federal education expenditures are substantial but as a percentage of total expenditures these appear small i.e., ranging between 1.7% - 2.3%.

Table 5.7: *Actual Education Expenditures against Total Expenditures (Rs. m)*

	2009/10			2010/11			2011/12		
	Edu Exp	**Total Prov Exp**	**%Edu/ Total Exp**	**Edu Exp**	**Total Prov Exp**	**%Edu/ Total Exp**	**Edu Exp**	**Total Prov Exp**	**%Edu/ Total Exp**
Punjab	106033	581373	18.2%	140168	668871	21.0%	166830	784650	21.3%
Sindh	52870	350672	15.1%	72395	428816	16.9%	47213	305138	15.5%
KP	39138	201141	19.5%	51030	254872	20.0%	67381	313965	21.5%
Baluchistan	7380	37617	19.6%	19409	128855	15.1%	23981	144474	16.6%
Federal	48482	2107885	2.3%	60488	2745271	2.2%	48151	2916263	1.7%
National	253903	3278688	7.7%	343483	4226685	8.1%	353556	4464490	7.9%

Source: Office of the Controller General, Accounts (CGA), 2012, Govt. of Pakistan

5.13.5 Distribution of Education Expenditures by Sub-Sectors

In 2011/12, national statistics show that pre and primary education expenditure was Rs 137,504 million i.e. the highest share (39%) in education expenditure, followed by Rs 119,642 million by secondary (34%) and Rs 74,935 million by tertiary (21%) sectors (Table 5.8).

Table 5.8: *National Education Expenditures by Sub-Sectors (Rs. m)*

Education Level / Type	2009/10	2010/11	2011/12	% (2011/12)
	Actual	**Actual**	**Actual**	
Pre and Primary Education	94652	129339	137504	39%
Secondary education	70219	103912	119642	34%
Tertiary Education	67715	85886	74935	21%
Education (undefined level)	1493	1734	2234	6%
Subsidiary Services	968	926	890	
Administration	10276	15049	7360	
Education (Others)	8581	6637	10992	
Total Education	253903	343483	353556	100%

Source: Office of the Controller General, Accounts (CGA), 2012, Govt. of Pakistan

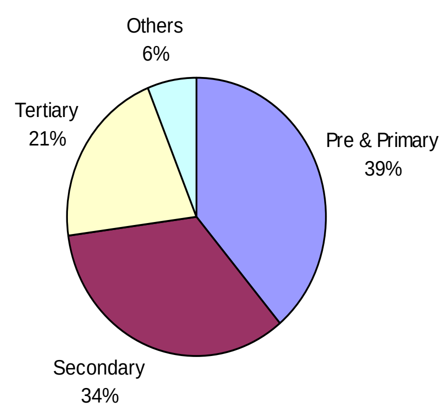


Figure 5.1: *Distribution of National Education Expenditures by Sub-Sectors*

On average at the national level, 95% of education expenditures comprise current expenses such as teachers' salaries (Table 5.8). Only 5% comprises development expenditures, which is not sufficient to raise quality of education. Across provinces, too, an overwhelming proportion of total actual education expenditures are spent on current heads, mainly teachers' salaries, leaving a small proportion for development expenditures. For 2011/12, except in Sindh where development expenditures are 16% of the total actual expenditures, these range between 1% (in Punjab) and 7% (in KP).

Table 5.9: *Education Expenditure Per Primary Student & Per Child*

	Actual Education Expenditure (Pre & Primary)* in Rs m	Total Public Primary Enrolment (#)	Education Expenditure Per Student (in Rs)
Punjab	75,371	4,207,122	17,915
Sindh**	25,205	2,197,866	11,468
KP	24,574	1,748,979	14,050
Baluchistan	7,683	466,523	16,469
Federal***	4,671	574,390	8,132
NATIONAL	137,504	9,194,880	14,954

* Education Expenditures are reported for Pre & Primary Education. As there are no specific expenses made on "katchi" it is assumed that most of this expenditure is made on primary education. ** not confirmed. *** includes ICT, FATA, GB and AJ&K

Source: Office of the Controller General, Accounts (CGA), 2012, Govt. of Pakistan; National Education Management Information System(NEMIS) Database 2011-12, AEPAM, MET, Islamabad; Population Projection 2005-2025, NIPS, 2013

5.14 Challenges to Education

In Pakistan, there are two key challenges in education:

1. Lack of access to education/high dropout rate; and
2. Poor quality of education

For each of these, a set of in-school and out-of-school factors were identified, along with some solutions for overcoming these challenges. Following is a matrix of these issues, along with a set of factors and corresponding strategies:

ACCESS	
In School Factors	***Strategies***
-Shortage of Teachers -Teacher absenteeism -Missing basic facilities -Lack of friendly environment -Teachers' harsh attitude	-Hiring of more teachers -Strict monitoring of teachers -Provision of basic facilities -No corporal punishment -Incentives to retain students e.g. free

	books, uniforms, food
Out of School Factors	***Strategies***
-Shortage of Schools -Distance especially for females -Insecurity -Poverty -Cultural norms -Parents are reluctant -Parents lack awareness -Law and Order Situation -Calamities	-More schools to be established -New schools established nearer to communities -NFE/Community/Feeder schools be established -Enforcement of security and law and order -Free education -Awareness campaigns on importance of education -Community involvement in education
QUALITY	
In School Factors	***Strategies***
-Ill Trained Teachers -Outdated Teaching methods -Missing facilities -Lack of Monitoring/Supervision	Training of Teachers/refresher courses Teaching methods to be revised Provision of basic facilities Involve community members to monitor schools
Out of School Factors	***Strategies***
-Poor Governance -Political pressures	-Effective monitoring and supervision -Rewards and punishments -Involve community members to monitor schools -End undue political interference and ensure merit-based appointments

5.14.1 Reasons for Slow Progress as per EFA/MDG Indicators in Pakistan

In Pakistan, one of the main factors of this slow progress in education indicators was a series of natural disasters, along with political events which affected the country during the past 7-8 years.

At the turn of this century, as countries began to understand and prepare for the implementation of the Millennium Development Goals (MDGs), Pakistan too initiated a process of rapid educational reforms. It launched the Education Sector Reforms (ESR) package and in some provinces, by 2002/03, the education sector reforms programs had been fully designed and even started to be implemented. Unfortunately, soon after, a strong earthquake in the northern part of the country left over 70,000 dead, millions homeless and a widespread destruction of schools, hospitals, roads and other infrastructure. As massive rehabilitation and reconstruction efforts took place, the progress towards MDG targets stalled.

The years 2007 and 2008 witnessed political instability and the transition from a military-led regime to a democratically-elected government also caused disruptions in economic and social development. This was coupled with the on-going militancy and extremism crisis in the north-west where military operations against the Taliban intensified. In July and August 2010, heavy monsoon rainfall caused flooding in the north and north-west, parts of Khyber Pakhtunakhwa (KP), Gilgit-Baltistan, Azad Jammu and Kashmir (AJ&K). As this large body of water made its way to south through the Indus River System, large areas of lands in Punjab and Sindh were inundated. These floods affected 78 districts and 20% of the country's area. A large number of schools were totally or partially damaged; remaining schools served as temporary shelters for the affected families.

In July 2011, the 18th Amendment to the Constitution of Pakistan became effective. This Amendment called for a transformation of government through devolution of power to the provinces. Education, too, was almost completely devolved. The federal Ministry of Education was dissolved and all decision-making powers given to the provinces. Education had always been a provincial subject in Pakistan but this formalized the withdrawal of federal coordination functions. As the bureaucratic systems began to adjust to the requirements of the new amendment, procedural delays in financial and technical issues adversary affected the education sector.

While reconstruction and rehabilitation of the 2010 flood affected areas was still underway, floods again hit some parts of country, particularly in Sindh and Baluchistan, in August 2011. Though the destruction was marginally lower than that in the previous year, over 9 million people were affected with huge loss of their assets.Once again schools and educational activities were adversely affected and progress in educational indicators slowed.

Given the high level of similarity between the EFA goals and education MDGs - in fact, EFA goals can be termed as a sub-set of MDGs (as MDGs represent a broader spectrum including poverty, employment, health, environment, etc.), the progress so far achieved in implementing the international commitments to education can be assessed by reviewing the change which may have taken place in a set of common indicators.

According to popular literature on the EFA[2] and MDGs[3], the set of common education indicators includes:

(a) Net primary enrolment rates - to reflect move towards Universal Primary Education;
(b) Completion/survival rates till grade V - as proxy for quality of education;
(c) Adult literacy rates;
(d) Gender Parity in primary education;
(e) Gender parity in secondary education;
(f) Youth (15-24 years old) literacy gender parity.

While experts are seriously concerned about Pakistan's probability of not meeting Education MDG targets, an effort has been made to demonstrate, on the basis of past achievements, to trace trend lines for various education MDG indicators. For this, two reliable sources of official data i.e., the Pakistan Social and Living Standards Measurement (PSLM) Survey and National Education Management Information System (NEMIS) have been used. The projections for education MDG indicators states that the net primary enrolment rate are likely to reach 78% by 2015/16 whereas given a declining trend in completion survival rates during 2006-07 to 2011-12 NEMIS data projects that survival rate will reach 65% by 2015-16.

The following table shows the summary of natural disasters along with political events which affected the country during the past 7-8 years:

Year	Disaster / Challenge	Outcome / Work Areas
2005	-Earthquake hit many areas of the country specially KP, FATA & AJK	-Many Schools destroyed -Enrolment decreased -Teachers strength decreased
2006	-Reconstruction and rehabilitation of earthquake affected areas	-Funds allocation -Involvement of international donor agencies -Earthquake reconstruction and rehabilitation agency (ERRA) established -Reconstruction of new schools -Rehabilitation of settlement of population affected by earthquake

[2] EFA Development Index (EDI) is based on UPE; Adult Literacy; Gender specific EDI; and Completion rates as proxy for quality of education

[3] MDG Reports typically base their analyses on net primary enrolment rates; completion/survival rates; adult literacy rates; and gender parity in primary & secondary enrolments and in youth literacy.

2007	-Disruptions in economic and social development -Militancy and extremism crisis in the KP province and FATA	-Many Schools destroyed by the terrorist activities -Settlement of Millions of Internally displaced persons -Mobile and feeder schools opened
2008	Transition from a military-led regime to a democratically-elected government	-Work on New Education Policy started
2009	-Heavy monsoon rainfall caused flooding in the north and north-west, parts of KP, GB, AJ&K -Large areas of lands in Punjab and Sindh inundated -Floods affected 78 districts and 20% of the country's area -Increase in terrorist activities in FATA	-A large number of schools were totally or partially damaged; -remaining schools served as temporary shelters for the affected families -New education policy announced -Many schools destroyed due to terrorist activities
2010		
2011	-Floods again hit some parts of country, particularly in Sindh and Baluchistan, in August 2011 -Over 9 million people were affected with huge loss of their assets. -July 2011, the 18^{th} Amendment to the Constitution of Pakistan became effective -devolution of power to the provinces -Education, too, was almost completely devolved -Terrorist activities and military operations carried on in FATA	-Schools and educational activities were adversely affected -Progress in educational indicators slowed. -Procedural delays in financial and technical issues adversary affected the education sector.

The details of the above mentioned issues are as follows:

Rationalization Policy in Punjab

After the General Election 2008, the Punjab Government started work on rationalization policy. According to this policy the institutions and teachers throughout the province were rationalized according to the needs of the respective area's population and enrolment. Many schools were closed and merged. A Large number of teachers were transferred to other areas. This policy produced some good results in areas but closure of nearby schools have negative impact on the enrolment. The following chart shows the decrease in number of institutions through the Punjab province in last four to five years.

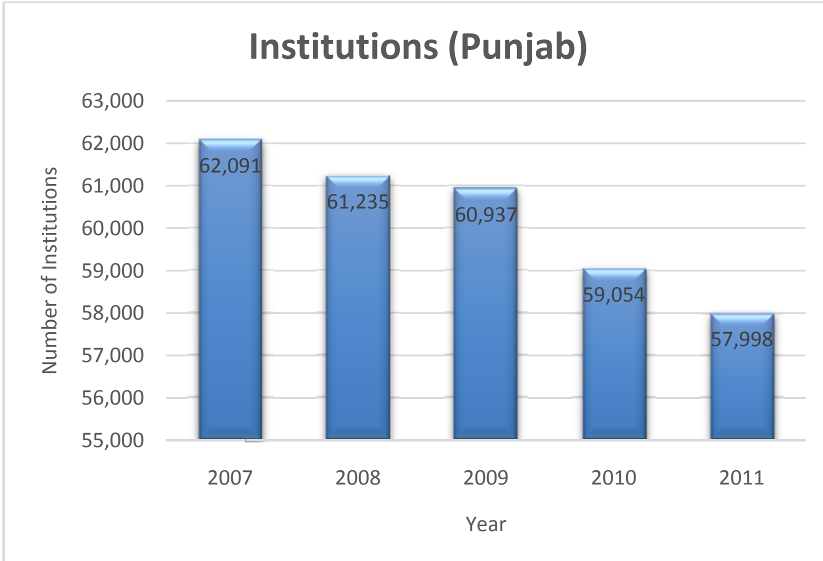


Chart 5.1:*Institutions of Province of Punjab*

Expenditure as Percentage of GDP

Pakistan's education expenditure as percentage of GDP has varied between 1.7% and 2.5%. Even with these small amounts, the utilization rates have remained at an average of 90%. The expenditure on education as percentage of GDP for the last many years have been depicted in the following chart:

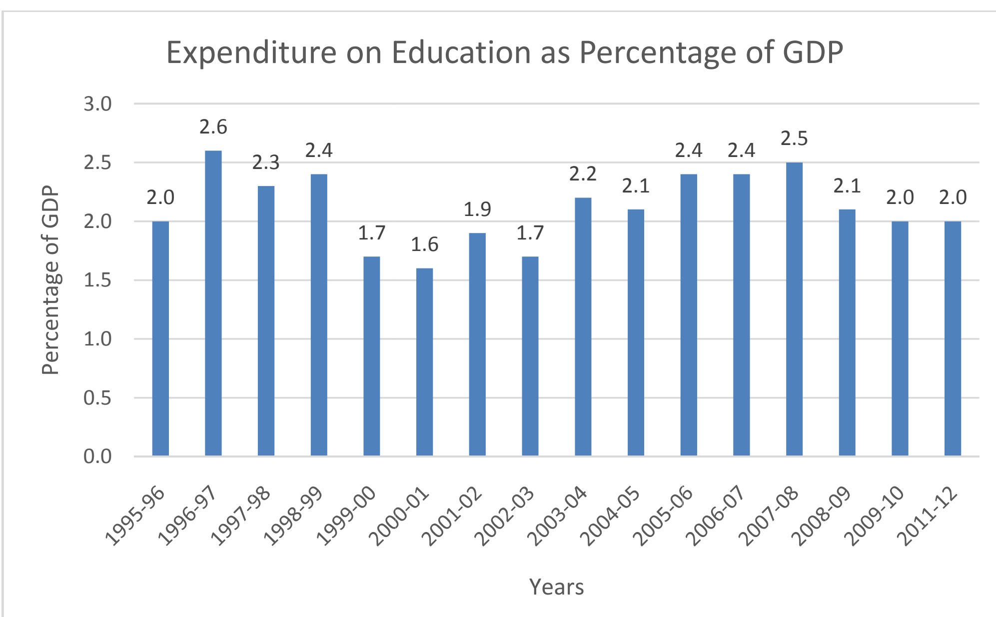


Chart 5.2:*Expenditure on Education as Percentage of GDP*

Law and Order Situation

The law and order situation in The FederallyAdministered Tribal Areas of Pakistan is very deteriorating. Many military operations were carried out in the Agencies and war on terror is also in progress. Many thousand people migrated to other areas of the country as Internally Displaced Persons. These issues has reduced the enrolment in the public sector schools. The following chart shows the decrease in enrolment due to these problems.

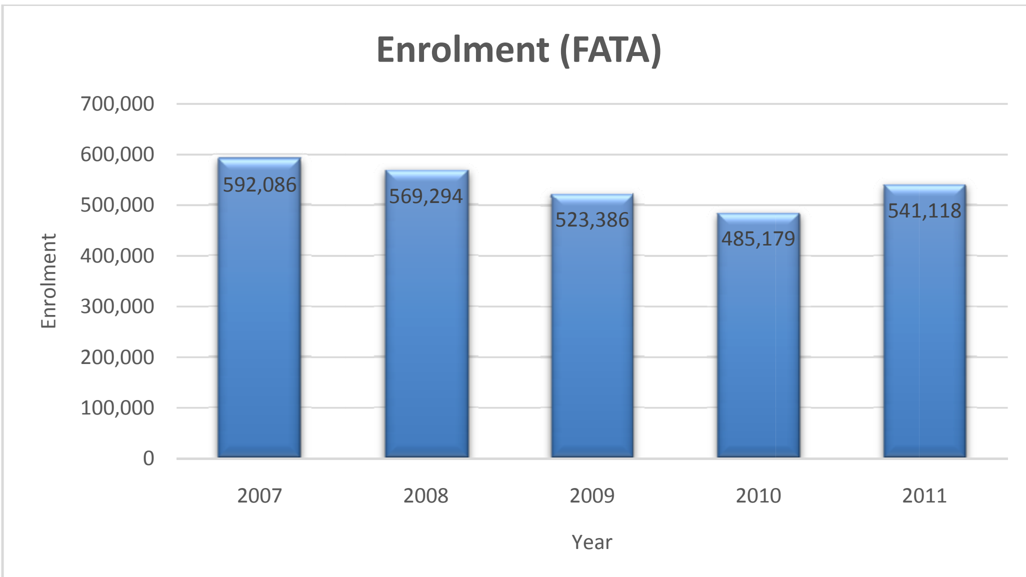


Chart 5.3:*Effect of Law and Order Situation on Enrolment in FATA*

Floods

In July and August 2010, heavy monsoon rainfall caused flooding in the north and north-west, parts of Khyber Pakhtunakhwa (KP). As this large body of water made its way to south through the Indus River System, large areas of lands in Punjab and Sindh were inundated. These floods affected 20% of the country's area. A large number of schools were totally or partially damaged; remaining schools served as temporary shelters for the affected families. While reconstruction and rehabilitation of the 2010 flood affected areas was still underway, floods again hit some parts of country, particularly in Sindh and Balochistan, in August 2011. Though the destruction was marginally lower than that in the previous year, over 9 million people were affected with huge loss of their assets. Once again schools and educational activities were adversely affected. The following Charts shows the effect of floods in number of institutions and enrolment:

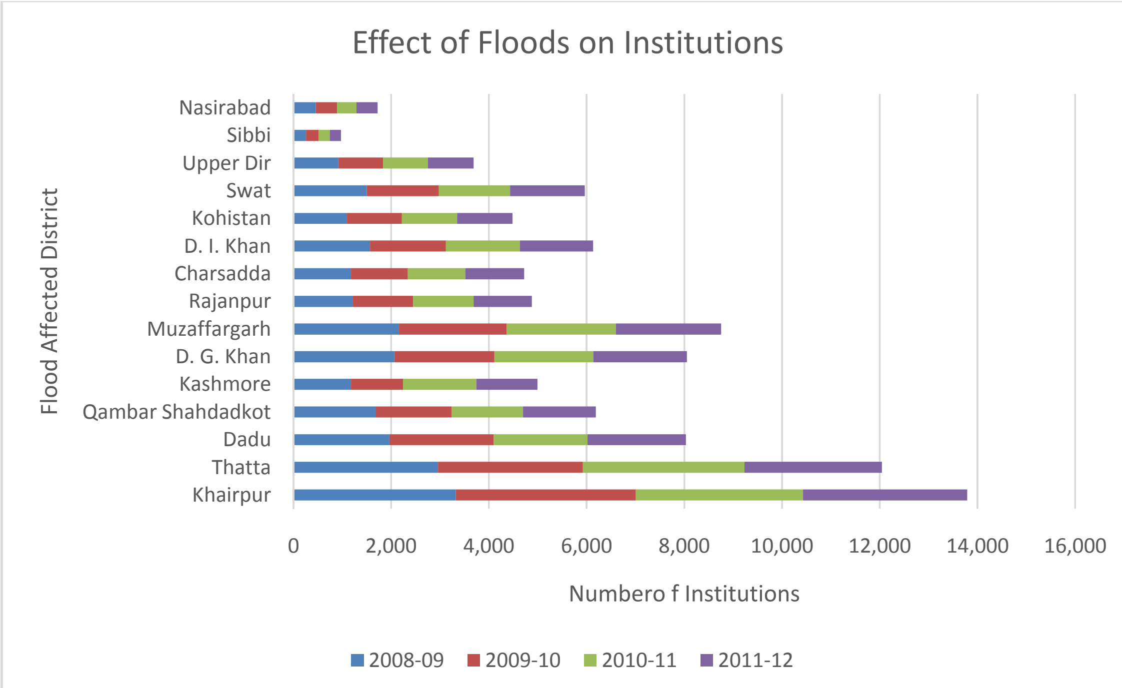


Chart 5.4:*Effect of Floods on Institutions*

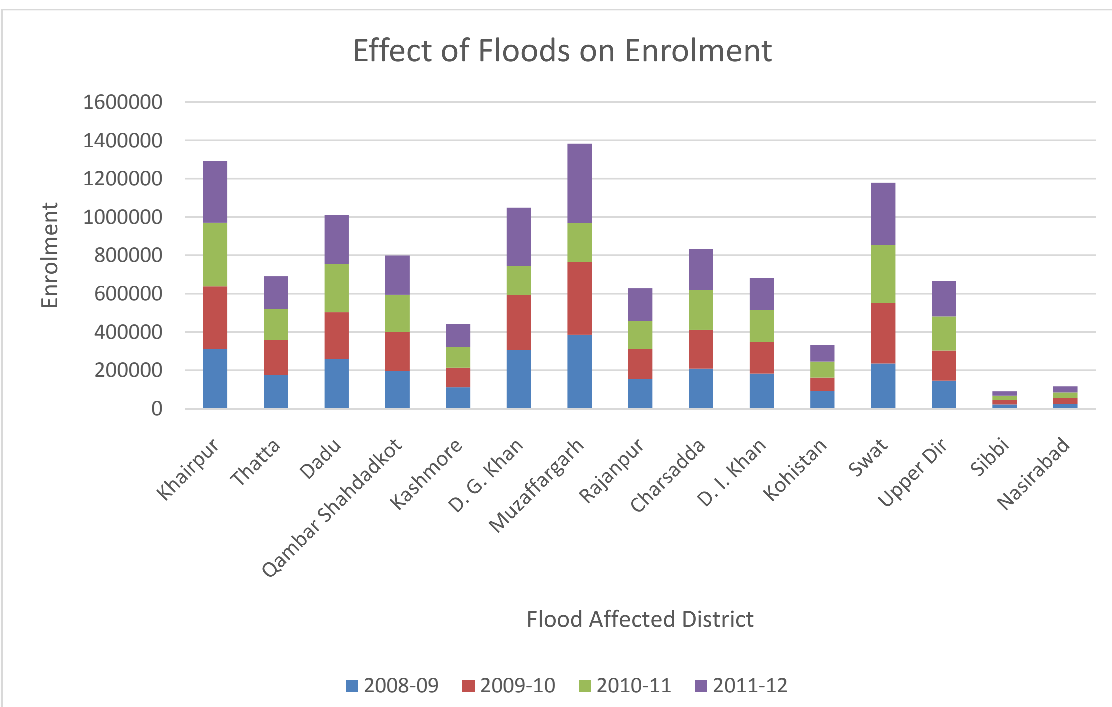


Chart 5.5:*Effect of Floods on Enrolment*

Chapter 6
CONCLUSION

6.1 Overview

The education is considered to be the fundamental right of every human being. The nations who intend to flourish invest heavily in their education system. Pakistan is among those countries which invest very less share of their GDP in education.

Education Management Information System is relatively new as compared to other fields of education. The significance of Education Management Information System cannot be denied as it provides the necessary baseline data that shows the current situation of education. This baseline data gains more important when the decisions are taken based on this data to set the direction of the education system. At present the EMISs in Pakistan lack the capacity of forecasting based on time-series analysis and other sophisticated data mining techniques. At present, the decision making and policy formulation is based on traditional manual projection systems which are not reliable. In order to overcome these shortcomings, the present study was held.

This study aimed in analyzing the time series data available at National Education Management Information System. This analysis was further used in producing various forecasts and indicators. These forecasts and indicators help the education managers to assess the performance of education system and make the necessary reforms.

This study has presented a Decision Support System tool namely EDSS that uses the time series analysis and forecasting techniques of data mining on the already data available with EMISs. The tool also has the capacity of calculating various indicators, mostly EFA and MDG, based on the forecasts done by EDSS. The other important feature of EDSS is the flexibility of comparing any number of districts and provinces at various levels of disaggregation. These forecasts are further supported by graphs and GIS based maps which helps the education managers in better understanding of the results.

Since the National Education management Information System has the responsibility of maintaining and disseminating the education statistics, therefore, the forecasts and indicators calculated by EDSS tool can also be provided to many national and international organization such as Economic Survey of Pakistan, PRSP, and UNESCO Institute of Statistics.

6.2 Future Work

Time Series Analysis and Forecasting are the fields of Data Mining which are merely been used in Education Management Information System. Since the Data Mining approaches has performed outclass in other fields of life, therefore, the research space is available for applying these techniques in the field of Education.

On the basis of present study, following are some of the areas that were identified in which sufficient research work is required:

- The Artificial Intelligence and Neural Networks techniques can be applied to the EMIS database in forecasting.
- The simulation models can be developed. The data from the EMIS database will serve as the baseline data for these simulation models. The targets of these simulation models can be obtained from the policy documents. These simulation models can be extended further by incorporating the flexibility of generating scenarios and their effects on education system.
- The GIS based systems are getting popularity day by day. Therefore the GIS based analysis and forecasting models in the field of education can be developed. These models can show graphically the weak areas which need priority in policy decisions.
- Since each province and region has its own EMIS, therefore the Integrated EMIS system can be developed. This model should be capable of running the same system at each provincial and regional EMIS. Thus the consistency in information and analysis can be obtained. This model can remove the time delays occur at various levels of data collection, entry and dissemination.

BIBLIOGRAPHY

[1] United Nations Universal Declaration of Human Rights 1948, United Nations.

[2] "Evolution of national educational management information system", Academy of Educational Planning and Management, Ministry of Education, Islamabad.

[3] "Development of education in Pakistan", Academy of Educational Planning and Management, Ministry of Education, Islamabad, 2006.

[4] K. Ahlawat, "Education management information system: Education decision support system", Paper presented in a workshop on education management information system/education decision support system, Cairo, Egypt, 2001.

[5] L. Carrizo, C. Sauvageot, N. Bella, "Education policies and strategies: Information tools for the preparation and monitoring of education plans", Paris, UNESCO, 2003.

[6] D.W. Chapman, "The role of education management information systems in improving educational quality. In Chapman & Carrier (Eds.), Improving the education quality: a global perspective", Greenwood Press, Westport, 1990.

[7] D.W. Chapman, "The rise and fall of an education management information system in Liberia", Journal of Education Policy 6(2), 133-143, 1991.

[8] D.W. Chapman, L.O. Mählck, "From data to action: Information systems in educational planning", IIEP, UNESCO, Paris, France, 1993.

[9] R. Ellison, "A practical guide to working with education management information systems", 2004.

[10] A. Gibson, M. Middleton, "Educational management information system design and prototype application as data cupboard", Proceedings of 17th Annual International Conference of Association of Management/International Association of Management, 17(2), (pp.145-151), 1999.

[11] H. Hua, J. Herstein, "Education management information system(EMIS): Integrated data and information systems and their implications in education management", Paper presented in the annual conference of comparative and international education society New Orleans, LA, 1993.

[12] S. S. Khandare, "Management information system", CBS Publishers, New Delhi, India, 2002

[13] V.A. Lacey, M.T. Ahmad, "The National education management information system in Pakistan: Problems and prospects", Ministry of Education, Islamabad, 1997.

[14] K. C. Laudon, J. P. Laudon, "Essentials of management information systems (3rd ed.)", Prentice Hall, Upper Saddle River, NJ, 1999.

[15] H. C. Lucas, "Information systems concepts for management (4th ed)", Mcgraw-Hill, Singapore, 1990.

[16] T. Lucey, "Management information systems (8th ed.)", DP Publication, London, UK, 1997.

[17] M. R. Malik, "Improving decision-making systems for decentralized primary education delivery in Pakistan", (Doctoral dissertation), Pardee RAND Graduate School, 2007.

[18] R. McHugh, "Strengthening education management information systems: better use of data for better strategic planning". In Puamau, P. & Teasdale, B. (Eds). Educational planning in the Pacific: Principals and guidelines (chap.7, p. 76-98), 2005.

[19] A. C. Mehta, "Successful experiments in strengthening educational management information system in India: DISE initiatives", National Institute for Educational Planning and Administration, India, 2005.

[20] K. Moses, "Information systems for education management", 2000.Retrieved from http://www.techknowlogia.org/TKL_active_pages2/SearchCurrent/main.asp?IssueNumber=19

[21] J. A. O'Brein, "Management information systems: Managing information technology in the e-business enterprise (5th ed.)", McGraw Hill, New York, 2002.

[22] D. Shah, "Decentralization in the education system of Pakistan: Policies and strategies", Paper presented in UNESCO Seminar on Decentralization Policies and Strategies in Education, Buenos Aires, Argentina, 2003.

[23] J. Shoobridge, A. Naqvi, "A review of present status of EMIS and education GIS in Pakistan: A summary report by the British Council for DFID and partners", The British Council, Islamabad, 2002.

[24] D. Shah, "Case Study of Educational Management Information System (EMIS) dynamics and decision making in Pakistan", (Doctoral dissertation, University of Education, Lahore), 2008.

[25] "Pakistan Education Statistics", 2010-11, Academy of Educational Planning and Management, Ministry of Education and Training, Islamabad.

[26] "National Education Policy 2009", Ministry of Education, Government of Pakistan, Islamabad, 2009.

[27] Government of Pakistan, Ministry of Education. (2008). Education for all: Mid-decade assessment (country report). Islamabad: Ministry of Education.

[28] M. Plummer, "Draft policies and guidelines for the selection of an EMIS solution", 2003. Retrieved on 6th June 2008 from http://www.oecs.org/oeru/documents/ICT/EMIS%20Selection%20Policy%20-%20Mar.%202003.pdf

[29] T. Cassidy, "Education management information systems (EMIS) development in Latin America and the Caribbean: Lessons and challenges", 2006. Retrieved from http://www.iadb.org/IDBDocs.cfm?docnum=741814.

[30] H. Khan, "Education management information system", Paper presented in national training course on educational management and supervision, Islamabad, Pakistan, 1998.

[31] S.F. Ahmad, T. M. Qureshi, "Education context in poverty reduction strategy paper II", Ministry of Education, Islamabad, Pakistan, 2007.

[32] T. D. LeBlanc, and K. D. LeBlanc, K.D., "Evaluation of and proposal to strengthen existing NEMIS, BEMIS, and SEMIS units and their support organization. Education Sector Reforms Assistance (ESRA) program", Islamabad, Pakistan, 2004.

[33] "Educational Policy and Strategy Simulation v2.1 User's Guide", UNESCO.

[34] Megan Banick , "Participatory Decision-Making with the Education Policy and Strategy Simulation (EPSSim) Model"

[35] "The EdStats Newsletter", Volume V Issue 5, April 2012, World Bank, Washington DC.

[36] "The EdStats Newsletter", Volume IV Issue 5, April 2011, World Bank, Washington DC.

[37] "The EdStats Newsletter", Volume IV Issue 4, February 2011, World Bank, Washington DC.

[38] "Principal features of Egypt-ANPRO-Model", IIEP Strategic Seminar, 19 February, 2009

[39] "Introduction to the ANalysis and PROjection Model - The ANPRO – Model", Mekong Institute-UNESCO course on Education Decentralization (13 February - 02 March 2007).

[40] "EPDC Enrolment Rates Model", Education Policy and Data Center, World Bank, 2007-2008. url: http://go.worldbank.org/TJNDSIEOH0

[41] Arun C. Mehta, "Projections of Population, Enrolment, and Teachers"

[42] E. G. Jacoby, "Methods of School Enrolment Projection", UNESCO, 1959.

[43] "Pakistan Millennium Development Goals Report-2010", Planning Commission of Pakistan, Government of Pakistan, Islamabad,

[44] "United Nations Millennium Declaration", Resolution adopted by the General Assembly, 8th plenary meeting, September 2000.

[45] "The Millennium Development Goals Report 2012", UNDP, United Nations, 2012.

[46] "World Declaration on Education for All and Framework for Action to meet basic learning needs", World Conference on Education for All, Jomtien, Thailand, March 1990.

[47] "World Declaration on Education For All", World Education Forum, 2000.

[48] "The Dakar Framework for Action", Adopted by the World Education Forum, Dakar, Senegal, 26-28 April 2000.

[49] "Global Education Digest 2012", UNESCO Institute of Statistics, United Nations, 2012.

[50] "The EFA Global Monitoring Report 2012", UNESCO, United Nations, 2012.

[51] Arun C. Mehta, "Indicators of Educational Development with focus on elementary education: Concept and Definitions",

[52] "Education Indicators Technical guidelines", UNESCO, November 2009.

[53] Dr. Pervez A. Shami, Dawood Shah , Mirza Tauhiduddin Ahmad, "National Education Core Indicators and Regional Comparison", Academy of Educational Planning and Management, Ministry of Education, Islamabad, August 2005

[54] "Reconstructed cohort method", UNESCO.
url: http://www4.unescobkk.org/education/efatraining/module-a3/annex-3/

[55] Dr. Hossein Arsham, "JavaScript E-labs learning objects for decision making", Merrick School of Business, University of Baltimore, Charles at Mount Royal, Baltimore, Maryland, 21201, USA.

[56] Merriam-Webster Dictionary

[57] Pam Perlich, "Extrapolation: Concepts and Techniques", University of Utah.

[58] http://en.wikipedia.org/wiki/Extrapolation

[59] http://en.wikipedia.org/wiki/Growth_curve

[60] M. Kahm, G. Hasenbrink, H. Lichtenberg-Fraté, J. Ludwig, M. Kschischo, "grofit: Fitting Biological Growth Curves with R". Journal of Statistical Software 33 (7). pp. 1–21, 2010.

[61] S. Makridakis, M. Hison, "ARMA models and the Box Jenkins Methodology"

[62] Box, Jenkins, Reinsel, "Time Series Analysis: Forecasting and Control", 1994.

[63] George Box, Gwilym Jenkins, "Time series analysis: Forecasting and control", San Francisco, 1970.

[64] Mario Jadrić , Željko Garača, Maja Ćukušić, "Student Dropout Analysis with Application of Data Mining Methods", Management, Vol. 15, 2010, 1, pp. 31-46, 2010.

[65] Pooja Gulati, Dr. Archana Sharma, "Educational Data Mining for Improving Educational Quality", International Journal of Computer Science and Information Technology & Security (IJCSITS), June 2012.

[66] Brijesh Kumar Baradwaj, Saurabh Pal, "Mining Educational Data to Analyze Students Performance", International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011

[67] Prof. Dr. P. K. Srimani, Malini M. Patil, "A Classification Model for Edu-Mining", International Conference on Intelligent Computational Systems (ICICS'2012), 2012.

[68] Tripti Arjariya, Shiv Kumar, Rakesh Shrivastava, Dinesh Varshney, "Data Mining and It's Approaches towards Higher Education Solutions", International Journal of Soft Computing and Engineering (IJSCE), Volume-1, Issue-5, November 2011.

[69] ShaeelaAyesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, "Data Mining Model for Higher Education System", European Journal of Scientific Research, Vol.43 No.1, 2010.

[70] Monika Goyal, Rajan Vohra, "Applications of Data Mining in Higher Education", International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012

[71] R. Lakshman Naik, S.S.V.N. Sarma, B. Manjula, D. Ramesh, "Study of Trends in Higher Education", International Journal of Computer Trends and Technology- Sep to Oct Issue 2011.

[72] Manpreet Singh Bhullar, Amritpal Kaur, "Use of Data Mining in Education Sector", Proceedings of the World Congress on Engineering and Computer Science 2012.

[73] Francisco J. Nogales, Javier Contreras, Antonio J. Conejo, Rosario Espínola`, "Forecasting Next-Day Electricity Prices by Time Series Models", IEEE transactions on Power Systems, Vol. 17, No. 2, May 2002.