2020-03-17

FACTOR IDENTIFICATION FOR QUALITY EDUCATION PERFORMANCE USING DATA MINING TECHNIQUE

BIRARA, ETSEHIWOT

BAHIR DAR UNIVERSITY

BAHIR DAR INSTITUTE OF TECHNOLOGY

SCHOOL OF RESEARCH AND POSTGRADUATE STUDIES

FACULTY OF COMPUTING

FACTOR IDENTIFICATION FOR QUALITY EDUCATION PERFORMANCE USING DATA MINING TECHNIQUES

BY

ETSEHIWOT BIRARA AYENEW

BAHIR DAR, ETHIOPIA

June 27, 2018

FACTOR IDENTIFICATION FOR QUALITY EDUCATION PERFORMANCE USING DATA MINING TECHNIQUE

ETSEHIWOT BIRARA

A thesis submitted to the school of Research and Graduate Studies of Bahir Dar Institute of Technology, BDU in partial fulfillment of the requirements for the degree of

Degree Of Master Of Science In Information Technology In The Faculty Of Computing

Advisor: Gebeyehu belay (PhD)

BahirDar,Ethiopia

June 27,2018

DECLARATION

I, the undersigned, declare that the thesis comprises my own work. In compliance with internationally accepted practices, I have acknowledged and refereed all materials used in this work. I understand that non-adherence to the principles of academic honesty and integrity, misrepresentation/ fabrication of any idea/data/fact/source will constitute sufficient ground for disciplinary action by the University and can also evoke penal action from the sources which have not been properly cited or acknowledged.

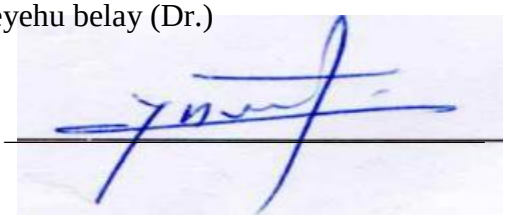
Name of the student Etsehiwot Birara Signature 

Date of submission: 11/09/2010 E.C

Place: Bahir Dar

This thesis has been submitted for examination with my approval as a university advisor.

Advisor Name: Gebeyehu belay (Dr.)

Advisor's Signature: 

Bahir Dar University

Bahir Dar Institute of Technology-

School of Research and Graduate Studies

Faculty of Computing

THESIS APPROVAL SHEET

Student:

Etsehiwot Birara _____ 11/09/2010E.C

Name Signature Date

The following graduate faculty members certify that this student has successfully presented the necessary written final thesis and oral presentation for partial fulfillment of the thesis requirements for the Degree of Master of Science in Information Technology

Approved By:

Advisor:

Gebeyehu B. (Dr.) _____ 11/07/2010.c

Name Signature Date

External Examiner:

Dr. Mesfin Kifle _____ 26/06/2018

Name Signature Date

Internal Examiner:

Mekuanint A. (PhD) _____ 16/07/2018

Name Signature Date

Chair Holder:

Bhabani Shankar D.M. _____ 23/07/2018

Name Signature Date

Faculty Dean:

Yemeriyin work _____ 11/09/2018GC

Name Signature Date

Faculty of Computing Program Manager

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr. Gebeyehu Belay for his continuous support in my MSc study and research, for his motivation, enthusiasm and patience. His guidance helped me in all the time of research and writing of this thesis.

I would like to thank amhara region education bureau education quality experts for their support and willingness to provide me the information I need. I would like to thank my friends for supporting me and encouraging me in every aspect.

Finally, I must express my very profound gratitude to my family; my dad Birara Ayenew, my mom Abeba Gela, my Sister and Brothers, for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

ABSTRACT

Education quality is the main concern of educational institutions and organizations. However, statistical data showed us that there is a big gap and challenges in quality education in terms of professional contribution to development, individual competence, and more specifically in students' performance. One of the solutions to reduce the problem of education quality is finding the causes and analyze through research using Data mining analytical tools. Therefore, the aim of this study is to identify education quality determinant factors. taking the domain 8520 Amhara region primary and secondary schools performance evaluation data of the years 2006-2008. we applied the J48 decision tree and JRip rule induction algorithms using WEKA data mining tool to build different models that identify the most determinant factors for education quality. After preprocessing a total of 8514 records are used for building the models and experiments are made to come up with a meaningful output. Major factors of quality education are identified and rules are generated using J48 decision trees and JRip rule induction algorithm with 84.67% and 84.80% accuracy respectively. The comparison of the models using WEKA's experimenter showed that JRip algorithm outperforms J48 algorithm. The most determinant factors for quality education identified by JRip algorithm includes: Teaching Learning Facilities, Financial Efficiency for Improvement, Student Participation, Teachers Education Delivery Performance, Students Sentiment, Responsibility and Behavior. Finally the finding of this thesis work could be a reference document for experts, decision makers and researchers who are interested in the field.

TABLE OF CONTENTS

Abbreviations

AI: Artificial Intelligence

ARFF: Attribute-Relation File Format

CRISP-DM KDP: CRoss-Industry Standard Process for Data Mining Knowledge Discovery Process

CSV: Comma Separated Value (Comma Delimited)

DM: Data Mining

EDM: Educational Data Mining

GNU General Public License

KDD: Knowledge Discovery in Database

KDP: Knowledge Discovery Process

MB: Mega Bits

MS-Excel: Microsoft- Excel

UNICEF: United Nations International Children Emergency Fund

WEKA: Waikato Environment for Knowledge Analysis

List of Figures

List if Tables

CHAPTER ONE

1 INTRODUCTION

Education, the process of facilitating wisdom or the acquirement of awareness, proficiency, ideals and attitude can be given in formal or informal way. Any practice that has influential consequence on one's thoughts, sentiment, or action is considered educational. Some of the general divisions of education are kindergarten, primary school, secondary school, college and university. A right to education has been recognized at the global level: Article 13 of the United Nations' 1966 International Covenant on Economic, Social and Cultural Rights.

As S.Visalaxi , S.Usha, S.Poonkuzhali (2015), all educational institutions strive to provide quality of education, which they stated that can be achieved by predicting the student's knowledge level of a particular subject. In Ethiopia different organizations and citizens are trying to assure "education for everyone". However it is now a fact that not only "education for everyone" but also quality education is the key to development. Therefore, quality education should be delivered for everyone to make students and graduates effective on their field. Ethiopian ministry of education formulated standards for education quality assurance in 2006e.c. Each school is evaluated according to the standard at least once in three years. Schools' which underperform or which are below the standard will be evaluated in consecutive two years. The major aim of education quality assurance standards is to assure education quality and effectiveness in the whole country to improve schools education delivery and students' education performance and behavior. The standard has three main parts named input, process and output. Each standard has its own sub standards with respective indicators. The input includes school facility, physical, human and financial resources. The process includes teaching and learning, curriculum, assessment, leadership, management and engagement with parents and the community. The output in education performance evaluation includes students' attainment and students' personal development.

Anchored in the hard effort of joint researchers, today's science is dependent to a great extent on data analysis. As the outlay for processing power and storage is diminishing data storage happens to be undemanding and economical, accordingly the amount of data accumulated in databases is escalating rapidly. Different Organizations accumulate enormous amount of data for years.

However, vast amount of these data sources create unfeasibility for human experts to come up with interesting information or patterns that will help in the practical decision making process. Manual data analysis technique has been touted as an alternative to analyze large data. However it is not capable of generating such valuable knowledge out of the unstructured data.

Data analytics is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making. Data mining is an advance data analysis tool that focuses on modeling and knowledge discovery for predictive purposes; uses computational techniques from statistics, machine learning and pattern recognition to analyze large data sets or databases. Data mining roots are traced along three family lines: Statistics, Artificial Intelligence (AI), and Machine Learning (cristina opera, 2014). Data Mining is one of the most motivating and vital area of research with the aim of extracting information from tremendous amount of accumulated data sets (Aboobyda J.H, Tarig M.A, 2016). It started to be an interest target for information industry, because of the existence of huge data containing large amounts of hidden knowledge (Qasem A. Al-Radaideh & Eman Al Nagi1, 2011). Data Mining is the advanced process which extracts the potential effective and comprehensive mode from the vast amounts of Data (Fang Weiping & Wang Yuming, 2013). Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data (Ryan S.J.d. Baker, 2010). It is a kind of technology, which combines the traditional data processing methods with different algorithms, to analyze new data types and extract knowledge from huge amounts of data (Fang Weiping Wang Yuming, 2013). The aim of data mining is to provide an automated data analysis. Apart from the need of getting useful knowledge, Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships needed for decision making (Pooja, Anil. & Manisha, 2015).

1.1 Statement of the Problem

Many researchers have studied the relationship and factors of students' performance and education quality by applying a data mining technology to come up with interesting information or patterns that will help in the decision making process. The researchers studied on different parameters of schools data.

In order to achieve Educational institutions goals, in Ethiopia it requires to apply data mining technology and explore interesting knowledge from the available databases. So that Schools can get directions on where to give more attention for the development of quality education by applying data mining technology to extract hidden knowledge from their available datasets. Educational institutions engage in diverse activities related to delivering quality education. Their primary task is to provide education for students, to make sure education is provided by well-educated and disciplined teachers and to provide necessary facilities for education. In addition to these activities, the fundamental task of educational institutions is to make sure quality education is being delivered for students.

One of the organizations that are trying to assure "education for everyone", for example, the Amhara region education bureau is now working towards delivering quality education for every citizen in the region. Even though Amhara region education bureau is working on delivering quality education, the schools current quality education delivery method is not efficient and effective. Not having efficient quality education delivery method makes the regional schools less competitive and contributes to a slow education development. Hence, it showed the need of applying advanced analytic tools and data mining technology to identify influential factors and improve education quality accordingly. These can be done using education quality performance evaluation standard presented by the ministry of education which has different parameters to evaluate schools performance.

Research questions:

- What are the factors that characterize quality education?
- What are the determinant variables for quality education?
- What are the most interesting patterns or rules generated using the determinant factors of quality education?
- How the application of data mining is supplementary for quality education evaluation?

Objectives

1.2.1 General objective

The general objective of the research is to analyze education quality performance and to identify the most determinate factor for the development of classification model.

1.2.2 Specific objectives

The research has the following specific objectives.

- ✓ To characterize the factors for quality education.
- ✓ To define the determinant factors for quality education.
- ✓ Augment data mining application for the improvement of education quality.

1.3 Scope

The scope of this thesis is, limited to the development of classification model by applying data mining technique. Even though the finding of the thesis is imperative for others in Ethiopia to a great extent, the scope of this thesis is restricted to amhara region primary and secondary schools because of time and resource limit.

1.4 Contribution of the Thesis

These days quality education is becoming the main concern of schools, governmental and nongovernmental organizations. Since Quality education is the key to development, primary and secondary schools education quality should be assured as they are the basis for all higher education levels. As being part of the society plus education experts, people and organizations working for education quality desires the development of education quality in amhara region. Keeping this in mind this research addresses the gap by identifying the determinants in education quality and allows them to attain useful information. Furthermore, they can be able to make decision according to the information. Whilst the performance evaluation has benefits for schools, Analyzing the schools performance data using data mining technique gives a great advantage for the quality education development. The factors behind the good or poor performance of schools can be recognized.

1.5 Organization of the Thesis

This thesis work is organized in five chapters.

The first Chapter briefly discusses introduction to the problem area, states the problem, the general and specific objectives of the study, the scope, and contribution of the thesis work.

Chapter two thoroughly discusses important literatures in the area of data mining technology and its application in the problem area as well as the technologies and methods that have been used and applied in other related research works.

Chapter three explains about the research methodologies used for the thesis work.

Chapter four is about analysis on results of the processing and discussion on the results.

Finally, chapter five provides conclusion of the research, and also presents recommendation for future work, which followed by the reference lists.

CHAPTER TWO
2 LITERATURE REVIEW

In the previous chapter we described the basic concepts of data mining. In this chapter we discused processes and algorithms in the context of data mining and educational datasets. We describe classification methods in data mining, and follow this with a study of the classification methods developed in related researches.

2.1 Quality Education Performance and Quality Education Characterization

Education, the process of facilitating learning, or the attainment of knowledge, skills, values, beliefs, and habits can take place in formal or informal situate. Any experience that has a formative effect on the way one thinks, feels, or acts may be considered educational. A right to education has been recognized at the global level: Article 13 of the United Nations' 1966 International Covenant on Economic, Social and Cultural Rights recognizes a universal right to education. By the start of the 21st century, the majority of all children in most regions of the world attended school.

Educational institutions which include preschools, primary schools, secondary-high schools, and universities where individuals acquire edification, offer a range of learning environments. The main motto of all educational institutions is to provide quality of education (S.Visalaxi , S.Usha, S.Poonkuzhali, 2015). Educational institutions help to direct the youth though the founding of a day to day practice, which has a greatest significance to direct them towards a better accomplishment, and as to make them turn out to be productive members of society.

According to UNICEF a paper “Defining Quality in Education” in June 2000, defines quality in the context of Education, which can be characterized by quality learners, learning environments, content, process and outcomes. The quality characterizations presented by UNICEF are explained below.

Quality learners: characterized by: students having

- Good health,
- Early childhood experiences and
- Home support.

Quality Learning Environments characterized by:

- Physical elements which include Quality of school facilities,
- Psychosocial elements which include Peaceful, safe environments, and the other are Service delivery which includes Provision of health services.

Quality Content: the intended and taught curriculum of schools including:

- Student-centered,
- Non-discriminatory,
- Standards-based curriculum structures.

Quality Processes characterized by:

- Teachers: including Professional learning for teachers and Teacher competence.
- Supervision and support which includes Administrative support and leadership.

Quality Outcomes characterized by:

- Formative assessment to improve achievement outcomes and
- Outcomes sought by parents.

As mentioned, Ethiopian ministry of education formulated standards for education quality assurance in 2006ec. Each school is evaluated according to the standard at least once in three years. Schools' which underperform or which are below the standard will be evaluated in consecutive two years. The major aim of education quality assurance standards is to assure education quality and effectiveness in the whole country in order to improve students' education performance and behavior(MOE, 2014). The standard can be characterized into three main categories named input, process and output. Each part has its own sub standards. For example, in Amhara region, there are eight thousand five hundred twenty primary and secondary schools. Each school has been evaluated according to the education performance evaluation standard in the year between 2006 and 2008ec. The evaluation result of schools has been used to level the schools regarding their performance. Schools get their result with respective feedback to improve their performance. However, until now, Ethiopian schools performance evaluation standards are not applied for any further analysis that needs to be useful for identifying the most determinant factors for quality education delivery.

2.2 Data Mining and Knowledge Discovery for Education

High volume of digital data is enabled through communication technologies like e-mail, internet, extranet; through enhancing technologies such as data warehousing, data mining and intelligent

technologies such as intelligent agents, internet search engines (Levin and Zahavi, 1999). Advances in communication technologies, computer hardware and database technologies have made it easier for organizations to collect, store and manipulate massive amounts of data.

growth of organizational databases in number and size through the availability of powerful and affordable database systems cause exponential increase in information resulting a demand for analyzing the voluminous data produced by organizations. Resulting the need for new techniques and tools that help to automatically identify patterns; transform the processed data to draw meaningful conclusions; and then extract knowledge from the rapidly growing volumes of digital data. Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in 0data that may be used to make valid predictions.

Various researches are being carried out in the area of data mining, an interdisciplinary field raised from database systems, data warehousing, statistics, machine learning, data visualization, information retrieval, neural networks, pattern recognition, spatial data analysis, and many more (Sumathi, 2006). The techniques of data mining can be applied to a wide variety of data repositories including databases, data warehouses, spatial data, multimedia data, Internet or web-based data and complex objects. Data mining can be applied in any area, including education, banking and insurance, medicine, security and communication.

2.2.1 Knowledge Discovery in Database (KDD)

KDD implies to the whole route of finding out useful knowledge from data, and Data Mining implies to a particular step in KDD process. Data mining is the use of precise algorithms for digging out patterns from data. KDD focuses on the whole process of knowledge discovery from data, including how the data are stored and accessed, how algorithms can be scaled to massive data sets and still run efficiently, how results can be interpreted and visualized, and how the overall man-machine interaction can usefully be modeled and supported. KDD process consists of an iterative sequence of steps. The additional steps in the KDD process, such as data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge, and proper interpretation of the result of mining are essential to ensure that useful knowledge is derived from the data. KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data (Fayyad, Piatesky-Shapiro, and Symth, 1996).

The KDD process involves using the database along with any required selection, preprocessing, sub sampling and transformations; applying data-mining methods to enumerate patterns from it; and evaluating the results of data mining to identify the subset of the enumerated patterns deemed knowledge. The data mining component of the KDD process is concerned with extracting patterns. The overall KDD process includes the evaluation and possible interpretation of the mined patterns to determine which patterns can be considered new knowledge. Thus we can say the overall process of building and implementing a data mining solution is referred to as KDD (Fayyad, Piatesky-Shapiro, and Symth, 1996).

2.2.2 Data Mining

The basic tasks of data mining and KD are to extract particular information from existing databases and convert it into understandable or sensible conclusions or knowledge. As indicated above, data mining can be viewed as a step in the KDD process that consists of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns (or models) over the data. Although data mining is a step in the knowledge discovery process, it is a more popular term than the longer term of knowledge discovery in databases in industry, in media and in database research environment (Sumathi, 2006). The authors also suggested adopting a broad view of data mining functionality: data mining is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories.

The Data Mining Process

According to the CRISP-DM KDP (CRoss-Industry Standard Process for Data Mining Knowledge Discovery Process) model, the six basic steps of data mining for knowledge discovery are:

- Business understanding
- Data understanding
- Data preparation
- Modeling
- Evaluation
- Deployment

Business understanding

The basic steps for data mining to knowledge discovery start with defining business problem. The precondition to knowledge discovery is to understand the problem area and the data in order for the algorithms to provide useful result and to identify the data mining tasks to be performed, prepare the data for mining, or correctly interpret the results. A clear statement of the objectives needs to be stated to make the best use of data mining. An understanding of the formulation and the potential of the data is necessary in order to understand what is included and what is not included in the data. The combination of these two activities drives the development of a goal for the work. The focus here is on creating understanding of objectives and requirements make a Data mining problem definition and design a preliminary project plan to achieve the objectives.

Data understanding

Data understanding, the initial data collection and familiarization have the aim of identifying data quality problems, creating initial insights into the data, and detection of interesting data subsets. The heterogeneous and wide variety data collected from data sources may not be appropriate for a particular analysis as it will be difficult to identify the most influential predictors for the data mining tools. Segmentation analysis or a data mining model can be applied to extract the relevant target set that participate in the data mining process.

Data preparation

The most time consuming task of the data mining process, preprocessing is performed before a model can be developed. The data stored in a database may reveal noise, exceptional cases, or incomplete data objects. When mining data regularities, these objects may confuse the process, causing the knowledge model constructed to over fit the data. As a result, the accuracy of the discovered patterns can be poor. In order to insure the accuracy of the data, cleaning, validation, and completion processes are performed to develop accurate database for data mining. Cleaning data refers to the process of reviewing the data to find incorrect characters or wrongly transmitted information. Data preprocessing also involves other data processing tasks such as overlaying of data from other resources, consolidating and amalgamating records, summarizing fields, checking for data integrity, detecting irregularities and illegal fields, filling in for missing values, trimming outliers, cleaning noise. While being tedious and somewhat boring, data preparation and

preprocessing is definitely a critical function of the knowledge discovery process with significant impact on the quality of the modeling results (Levin & Zahavi, 1999).

The next important step of data mining process is data reduction and projection. This means finding useful features to represent the data depending on the goal of the task. The goal of exploring the data is to identify the most important fields in predicting an outcome, and determine which derived values may be useful. This step is the final data preparation step before building models. This step covers all activities needed to construct the final dataset, which constitutes the data that will be fed into DM tools in the next step. It includes Table, record, and attribute selection; data cleaning; construction of new attributes; and transformation of data.

Modeling

It is only at this point that one invokes data mining models and tools to interrogate the data and convert it into knowledge for decision making. At this stage, a particular data-mining method is selected that matches the goals of the data mining process defined in the first step. As mentioned earlier, the data mining component of the KDD process often involves repeated iterative application of particular data mining methods in searching for patterns of interest.

Diverse modeling techniques are selected and applied at this stage. Modeling usually entails the utilization of numerous methods for the same DM problem and the calibration of their parameters to optimal values. Since some methods may require a specific format for input data, often reiteration into the previous step is necessary. This step is subdivided into selection of modeling techniques, generation of test design, creation of models, and assessment of generated models.

Evaluation

Modeling provides a set of outputs which needs to be evaluated and interpreted to confirm the resulting model is good quality, and to convert the model results into useful knowledge for decision making. Taking the results of the data mining models for granted, without evaluation process could be very risky. Knowledge evaluation is often conducted by means of statistical measures and tools. After one or more models have been built that have high quality from a data analysis perspective, the model is evaluated from a business objective perspective. A review of the steps executed to construct the model is also performed.

Deployment

Finally the discovered knowledge should be expressed in expressive knowledge representation techniques, such as trees, tables, rules, graphs, and charts, cross tabs, matrices or other expressive forms so that the knowledge can be easily understood and be directly usable by humans. And must be organized and presented in a way that the customer can use.

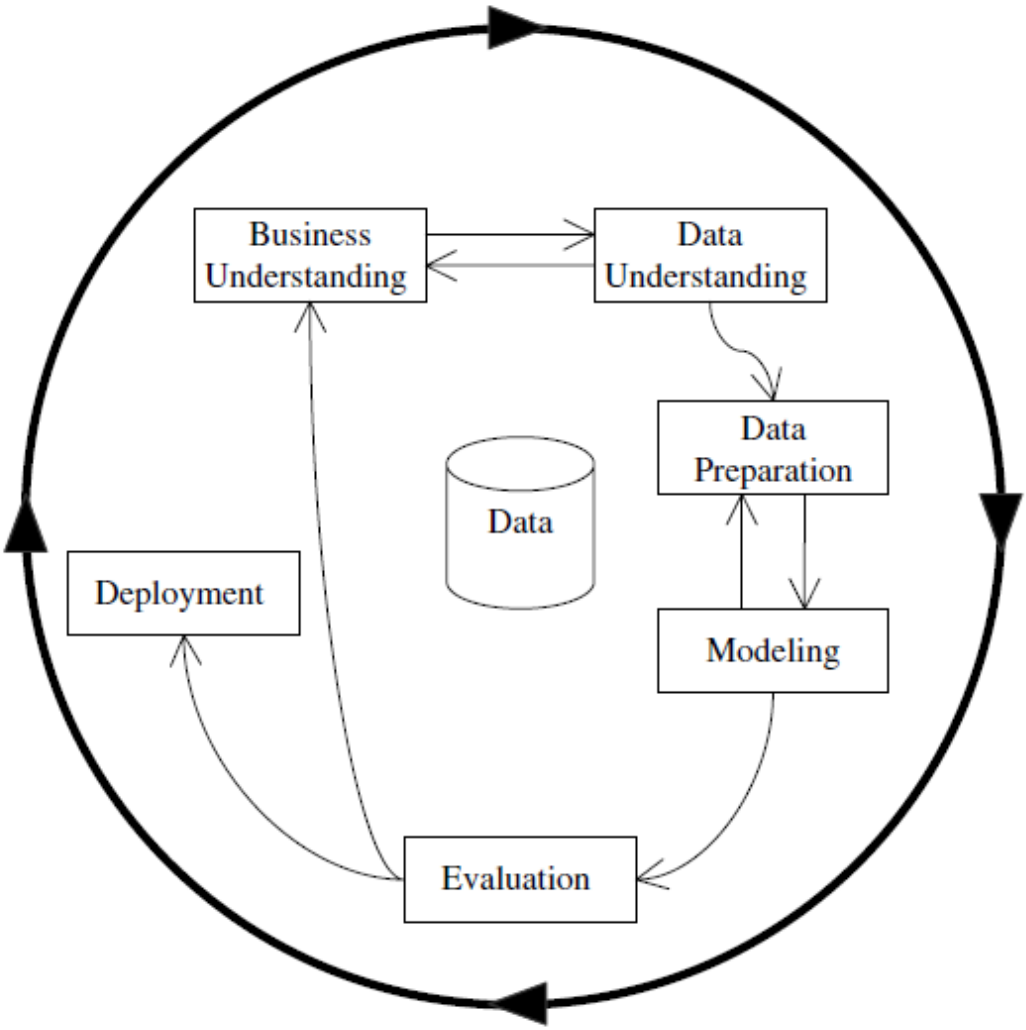


Figure 1: The CRISP-DM KD process model (source: http://www.crisp-dm.org/)

Data Mining Tool for Education Data Processing

The Analysis process of a study is highly influenced by the selected tool. In order to get best analysis results it is necessary to appraise suitable tools and techniques early in the process.

WEKA, being data mining software can be chosen for this particular education data set analysis study for the following reasons. WEKA data mining software can run on Windows as well as Linux operating systems and have the ability to access MS-Excel data formats supporting up to 65,000 records with memory expansion; it also provides classification features on the data sets and visualizes the output of the processes. The above mentioned characteristics of WEKA leads the researcher to choose it for the study as the researcher uses Windows operating system and the education data is in MS-Excel data format having 39 parameters and 8520 records. Since the study focuses on classification of education data and analyzing the result of the classification by visualizing the output it is appropriate to select and to use WEKA data mining software. In addition to, the above mentioned reasons other reasons that make the researcher to choose WEKA as a data mining tool for this particular study are it is open source software, easily accessible and provides different functionalities such as classification, clustering, association, attribute selection, and visualization.

WEKA

The freely available software WEKA (Waikato Environment for Knowledge Analysis) which is implemented in Java programming language under the General Public License (GNU) is free software that provides a diversity of machine learning algorithms and tools having features used for preprocessing and processing capabilities in data mining tasks. WEKA offers a simplified data preprocessing and processing abilities through organizing, fixing missing value, cleaning noisy data, clustering, classification, association mining algorithms, attribute selection and data visualization. Different Data file formats can be executed in WEKA including ARFF, CSV, XML and spreadsheet. The ARFF is widely used for analysis of datasets. WEKA also has capability to provide access to datasets collected in DB2, MySQL, MS Access, or other formats through JDBC. Although Weka provides the above mentioned features for data mining it does not have the multi relational data mining capability (An Introduction to WEKA, PACE).

Data preprocessing

Data preprocessing, preparing data to be used for analysis purpose is the first step to be applied in WEKA before performing data analysis. It is a necessary step in data mining since most data collection methods are not accurate. A gathering of diverse methods, Data preprocessing is considered to be applied prior to the actual analysis in order to formulate the collected data in a way

that is suitable for analysis. Data cleaning, Data reduction, Data discretization, data selection, Data integration and Data transformation are the key Tasks in Data Preprocessing.

As mentioned above in the data collection section the data used for this particular research work is particularly a secondary data that is collected from each Amhara region school by the regional education bureau. The data have need of a bunch of amendment as it is secondary data so the preprocessing tasks are needed to be applied. The fact, quality decisions are based on quality data indicates that no quality data, no quality mining results.

Data cleaning, task of preprocessing is the process of making the dataset suitable for analysis by diminishing introduction of foregone conclusion into the dataset. Datasets especially those that are collected through manual data collection methods such as interview, questioner and observation are prone to errors like outliers, missing values of records, noisy data, and duplicate records and so on.

An incomplete or missing value is characterized by lacking attribute values, lacking certain attributes of interest, or containing only aggregate data due to data not entered, equipment fault, deleted due to inconsistency and so on. In Data cleaning preprocessing task missing data values are handled by applying a number of different procedures. The most frequently used data cleaning techniques are ignoring the data row, using a global constant to fill in for missing values, Using attribute mean, using a data mining algorithm to predict the most probable value.

Noisy data can be characterized by data that contains errors or outliers due to faulty data collection, data entry, data diffusion methods. In Data cleaning preprocessing task noisy data are handled by applying Binning method, Clustering, Regression, Inconsistent data problems are characterized by data containing difference in codes or names due to inconsistency in naming convention. Inconsistent data problem can be solved by using fixed standards for coding and naming.

Data reduction, task of preprocessing is process of reducing the representation of the dataset in volume by maintaining the integrity of the original data for efficient mining through applying different strategies such as Dimensionality reduction, Data Reduction through sampling and Data compression strategies. Data discretization, Part of data reduction having particular importance, for numerical data reduction.

Attribute selection, task of preprocessing that returns a subset of the features is process of selecting significant attributes, reducing and removing redundant and irrelevant input attributes to identify important parameters that are necessary for processing and analysis purpose as redundant data makes it more difficult to discover meaningful patterns.

Data integration, task of preprocessing is process of combining data from multiple sources, integrating multiple databases, data cubes, or files in order to have a single integrated dataset that is suitable for data analysis purpose.

Data transformation, task of preprocessing is process of converting data or information from one format to another, usually from the format of a source system into the required format of a new destination system through Normalization and aggregation to improve the quality of input data as to produce the better quality results.

2.3 Implementation approaches

Applying the computing process for discovering patterns in large data sets, extracting information from a data set and transforming it into an understandable structure for further use is the main concern of this research work. In order to achieve the above mentioned task data mining approach is applied which is very popular in exploring hidden patterns and relationships in datasets that can be used to make predictions.

As mentioned above, the Data Mining (DM) conception is to dig out hidden patterns and to determine relationships among factors in a vast amount of data using data analysis tools. (Umesh Kumar Pandey S. Pal.) Data mining techniques that are used for mining and data discovery operation can be grouped based on their use as predictive or descriptive. Prediction involves using some variables or fields in the data base to predict unknown or future values of other variables of interest. Predictive mining tasks perform deduction on the current data in order to make prediction. In predictive models, the values or classes to be predicted are called the response, dependent or target variables. The values used to make the prediction are called the predictor or independent variables. Predictive models are built, using data for which the value of the response variable is already known which makes prediction to be considered as supervised learning because calculated or estimated

values are compared with the known results. The predictive analysis can take place through different techniques Classification, Regression, Time series Analysis.

Descriptive mining focuses on finding human-interpretable patterns (Fayyad, 1996). In Descriptive techniques, there is no already known result to guide the algorithms. Descriptive analytics looks at past performance data, it understands and analyzes past events and use data to look for the reasons behind past success or failure and to characterize the general properties of data. Clustering, Summarization, Association Rules, Sequence Discovery are descriptive analysis techniques.

2.3.1 Data classification technique

Classification technique, The best & popular predictive data mining technique makes Predictions on discrete variables, based on other attributes in the dataset or using known results found from different data by finding a set of models or functions that describe and distinguish data classes or concepts(Qasem A. Eman Al Nagi, 2011).

Classification, supervised learning technique which classify data item into predefined class label build classification models from an input data set that are used to predict future data trends. Classification is a two-step process, the learning step and the testing step, building a model describing a predetermined set of classes by analyzing a set of training database instances and testing the model using a different data set that is used to estimate the classification accuracy of the model respectively. If the accuracy of the model is considered acceptable, the model can be used to classify future data instances for which the class label is not known as the goal of classification is to accurately predict the target class for each case in the data. At the end, the model acts as a classifier in the decision making process. There are several data mining algorithms that can be used for classification such as decision tree, Bayesian methods, rule based, and Neural Networks.

The data mining algorithms analyze the data in hand and search explicit patterns in it (S. Sumathi, S.N. Sivanandam, 2006). The algorithm uses the results of analysis over much iteration to find the optimal parameters for creating the mining model. Different algorithms can be used to perform the same analytic task though the result produced by each algorithm is different, some algorithms can also produce more than one type of result. It is a challenging task to choose the best algorithm to use

for a specific analytical task. Here discussed, some of the popular data mining classification algorithms used in this particular research work.

Decision tree algorithm

Decision Tree classification algorithm, a flowchart-like tree structure where each internal node denotes a test on an attribute. Each branch represents an outcome of the test, and each leaf node holds a class label. Decision trees are most popular data mining techniques used for classification(Himani S., Sunil K., 2016). A decision support tool makes use of a tree-like graph of decisions and their corresponding consequences. It applies on categorical and numerical data for predictive modeling of discrete and continuous attributes by building a data mining model through creating a series of splits in the tree, splits representing nodes added to the model every time an input column is found to be significantly correlated with the predictable column at the same time an associated decision tree being incrementally developed. The end result is a tree with decision nodes and leaf nodes representing the best predictor root node and a decision respectively with the goal to create a model predicting the value of a target variable based on several input variables. The construction of decision tree classifiers does not require any domain knowledge or parameter setting. The learning and classification steps of decision tree induction are simple and fast with good accuracy.

J48 Algorithm

J48 is an improved version of an earlier algorithm developed by J. Ross Quinlan, C4.5. Decision trees, offer a fast and powerful way to express structures in data based on the attribute values of the available training data in order to classify a new item. Whenever it encounters a training set it identifies the attribute that discriminates the various instances most clearly giving more information about the data instances so that they can be classified. J48 constructs a decision node using the expected values of the class. J48 decision tree can handle specific characteristics, lost or missing attribute values of the data and differing attribute costs while precision can be increased by pruning.

Rule based classification Algorithm

A series of if-then statements that utilizes a set of assertions, to which rules are created on how to act upon those assertions. Rule based systems often construct the basis of software artifacts which can provide answers to problems in place of human experts. Provide an

adaptable method, suitable for a number of Different problems. Rule based systems are appropriate for fields, where the problem area can be written in the form of if-Then rule statements.some of the rule based algorithms are JRip, PART, decision table,ZeroR,OneR.

2.4 Data mining for educational data

Data mining is being applied worldwide in education areas to evaluate schools and students performances, and to improve their future performance. Diversity of data analysis tools are used for pattern and relationship determination.

Data mining assists analysts to find patterns and relationships in data. The quality of the output from the data mining process will often be sensitive to outliers and irrelevant columns depending on the way the data is encoded. Although a good data mining tool makes the work of the analyst easy by taking care of the details of statistical techniques, it requires understanding the workings of the tools that are chosen and the algorithms on which they are based. Data mining does not replace skilled analysts, but rather gives them a powerful new tool to improve the job they are doing (S.Visalaxi , S.Usha, S.Poonkuzhali april 2015).

Data mining techniques plays a vital role in all the application areas of education research and development. In the current times, one of the biggest challenges that educational institutions face both in the developing and developed nations is the explosive growth of educational data and to use this data to improve the quality of managerial decisions (Mohammed I. Al-Twijri, Amin Y. Noaman, 2015). Data mining aims to provide an automated data analysis. Apart from the need of getting useful knowledge, Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships needed for decision making (Fadhilah Ahmad, Nur Hafieza Ismail and Azwa Abdul Aziz 2015). Data mining techniques, which extract information from huge amount of data, have been becoming popular in education domains (Sumit Garg& Arvind K. Sharma 2015). Different researches have been done on student performance evaluation, improvement and their future performance.

2.5 Educational Data Mining

Educational data mining (EDM), a promising discipline concerned with developing scheme in support of discovering unique type of data from educational settings. EDM includes psychological

and computational schemes plus research approaches. Its crucial point is to apply data mining tools and techniques to any type of educationally related data (Ms. Priti S. Patel, Dr. S.G.Desai). Data can be students' data that is related with their class performance, exam scores, and behavior, also data may be academic staff's performance and ability or it can be about the schools education delivery performances and their ability to deliver education in appropriate and suitable way.

Educational data mining (EDM) is an interesting research area which extracts useful, previously unknown patterns from educational database for better understanding, improved educational performance and assessment of the student learning process (M. Durairaj, C. Vijitha 2014).

Educational Data Mining, Mining in educational environment concern with developing new methods to discover knowledge from educational database (Priyanka Saini Banasthali Vidyapith, Rajasthan). EDM deals with the application of data mining tools and techniques to inspect the data at educational institutions for deriving knowledge. The primary goal of EDM is to use large-scale educational data sets to better understand learning and to provide information about the learning process. EDM explores the row data from educational environment to useful data that can be used to making decisions and solve the problems.

Data mining is widely used in educational field to find the problems that arise in the field. Schools performance is of great concern where several factors may affect the performance. For prediction the three required components are: Parameters which affect the student performance, Data mining methods and the data mining tool. knowledge can be obtained by applying data mining techniques on schools data, which describes the school performance to improve education quality and to decrease failure rate of students. This study is conducted to maintain education quality of schools by identifying the most determinant factors.

In recent years, there has been increasing interest in the use of DM to investigate educational field. Educational Data Mining (EDM) is concerned with developing methods and analyzing educational content to enable better understanding of students' performance (Mohammed I. Al-Twijri, Amin Y. Noaman 2015). It is also important to enhance teaching and learning process. The data can be collected from historical and operational data that reside in the databases of educational institutes. The schools data can also be collected from schools by collecting e-learning systems which have a

large amount of information used by most institutes. Educational Data Mining uses many techniques such as Decision Trees, Neural Networks, Naïve Bayes, K-Nearest neighbor, and many others.

As EDM focuses on discovering new patterns which can be best applied to predict schools and student performance and to predict students' future performance, to identify the abilities of students, their interests and weaknesses, to identify students which need special attention, to cluster the students according to their performance and to assist students to improve their results by conducting a comparative analysis (Gurmeet Kaur, Williamjit Singh 2016).

EDM develops methods and applies techniques from statistics, machine learning, and data mining to analyze data collected during teaching and learning. EDM tests learning theories and informs educational practice. Various classifiers i.e Random Tree classifier, J48, REP Tree (ID3), Naïve Bayes and IBK algorithm can be used to predict Student knowledge level and to identify the learning behavior of students and the potential knowledge level of student (S.Visalaxi, S.Usha, S.Poonkuzhali 2015).

2.6 Litrature Survay

Fadhilah Ahmad, Nur Hafieza Ismail and Azwa Abdul Aziz in their work, Prediction of Students' Academic Performance Using Classification Data Mining Techniques they Produce students' academic performance prediction model using Students' demographics, previous academic records, and family background information which is Collected from 8 year period. they provide a Framework for predicting students' academic performance of first year bachelor students using Decision Tree, Naïve Bayes, and Rule Based And selected Rule Based algorithm as best algorithm with 71.3% accuracy.

Gurmeet kaur, williamjit singh in their work, Prediction of student performance using weka tool they Used 52 students data records to Maintain the education quality of institute. They used Naïve bayes and j48 decision tree classification techniques to Predict student performance. Naïve bayes algorithm outperforms j48 algorithm with accuracy 63.59 % and 61.53% respectively.

Umesh Kumar Pandey S. Pal, Data Mining in their work, A prediction of performer or underperformer using classification they Used 600 students data records of the year 2009-2010 to

Predict the students division on the basis of previous year database using Byes classification on Matlab environment.

Aboobyda Jafar Hamid and Tarig Mohammed Ahmed: Developing prediction model of loan Risk in banks using data mining they Used a collection of data from banking sector to provide new model for classifying loan risk in banking sector using J48, bayesnet and naivebayes algorithms. The accuracy of the algorithms shows J48 algorithm outperforms other algorithms with 78.3784 % accuracy, bayesnet 77.4775 %, naivebayes 73.8739 % accuracy.

Muhammed Salman Shamsi and Jhansi Lakshmi: Student performance prediction using classification data mining techniques. they used Students data including , academics, socio-economic state and Psychological state to Predict The performance of students in terms of grades and dropout for a Subject using Naïve Bayes, J48, random forest, and jrip algorithms on WEKA enviroment. The outputs of the experments shows that Jrip algorithm is more accurate than the other algorithms.

CHAPTER THREE
3 RESEARCH METHODOLOGY

This chapter provides information about the pre analysis activities that are performed in order to have effective analysis process. These include how the data collection process is performed, the data understanding along with the business understanding, the data formatting and the data preparation process for analysis.

3.1 Data Collection

Data mining exist with data. Analyzing and evaluating schools performance requires application of different methodologies. The first requirement for schools performance analysis is having schools data in hand.

The data source for this research consumption is the education quality performance of schools in Amhara region which is available in an Excel data format. The data includes primary and secondary schools performance evaluation records of the years 2006-2008 from 8520 schools. Since Collecting Data from 8520 schools is not possible because of time and resource limitations, we use secondary the data that is available in the regional education bureau.

The primary and secondary schools data collected from each amhara region primary and secondary school include information about the schools, such as school code, type, wereda, zone, location, ownership and other evaluation standards. The records in the data set specifies the values given for each school according to the performance evaluation standards in addition to the information about schools. These includes the schools physical, human and financial resources, teachers education delivery performance and students improvement and development. In addition to this, the grade which is calculated from the results of each performance evaluation of the school specifies the performance of the school by giving values 1, 2, 3, or ,4 which indicates poor, on improvement, good and high performance respectively.

3.1.1 Data Preparation

One of the most important tasks in data mining is preparing the data in a way that is suitable for the specific mining techniques, algorithms or software package to be used. Data preparation involves data selection, data cleaning, data construction, data integration and data formatting processes (The CRISP-DM, 2000).

In any data mining task the first step is clear understanding of the problem to be solved. In this chapter data understanding activities such as; data collection, data description and data formatting have been undertaken. Secondly, the data has been pre-processed by employing data cleaning and data selection techniques.

3.1.2 Data Understanding

The data understanding phase we start with an initial data collection and proceed with consulting the domain experts and reviewed related documents to create a clear insight and understanding towards the data and the problem domain.

Business process understanding

In addition to consultancy of domain experts and reviewing related documents, knowing the business process also help us to have a good understanding of the data.

The data used for this research work is gathered by education evaluation stafs of the regional education bureau from 8520 amhara region elementary and secondary schools to be used for schools performance evaluation. The data includes information about each school such as Zone, Woreda, location, code, name, type, ownership, year. Other parameters in the dataset include different evaluation standards such as, Facilities, Financial Efficiency, Staff efficiency, Environment, mission, plans, student participation, Students improvement, Students sentiment, Teachers performance, Modern teaching methods, Curriculum evaluation, Students evaluation, controls over plans, standards for resources, School parent relationship, development goals, Students result improvement, Responsible students, Staffs sense of responsibility, Strong relationship and support of parents.

As stated earlier, schools' education quality performance evaluation is performed by education quality evaluation experts from the regional education bureau and from the respective Woredas.

Education quality experts observe the schools for consecutive days (4-6 days) interviewed the school society and students and fill the evaluation form according to the school performance. Once the evaluation form is filled the rank is calculated based on the values given for each standard. Amhara region education bureau generates report about education quality of the evaluated schools after Collecting the evaluated schools data.

3.1.3 Formatting the Data

Conducting data mining on a collection of stored datasets requires various steps. Before conducting data mining the data stored must be prepared or preprocessed in an appropriate format that can be used in the data mining process.

WEKA data mining tool requires the dataset to be in a comma separated file format called the Attribute Relation File Format (ARFF). The ARFF file format is the standard way of representing datasets that consist of independent, unordered instances and does not involve relationships between instances. Hence, the education data set which was originally in an Excel file format is converted to an ARFF file format.

3.1.4 Data Description

Our data, a record of Amhara region schools performance evaluation is 10.2 MB (10,745,533 bytes) with 8,520 instances and 39 attributes of text and number types. Some of the attributes of the dataset are shown in table 1.

Table 1: Description of the attributes

No.	Attribute	Type	Description
1	Name	Text	Name of the inspected school
2	Zone	Text	Zone where the school is located
3	Woreda	Text	Woreda where the school is located
4	Code	Number	Unique identifier for each school
5	Type	Text	Primary/ secondary school
6	Ownership	Text	Private/ governmental school

7	Location	Text	Rural/ urban school
8	Year	Number	Inspection year
9	TeachingLearningFacilities	Number	The fulfillment of the necessary teaching learning facilities and buildings, supporting materials in the school.
10	FinancialEfficiencyforImprovement	Number	The school's financial efficiency for primary improvements in teaching learning process.
11	StaffEfficiency	Number	The school has efficient dean, teachers, and support staffs with appropriate education level and behavior.
12	SafeTeachingLearningEnvironment	Number	The school has provided safe teaching learning environment for the school society.
13	DevelopmentSociety	Number	the school has created education society that understands and is willing to practice the schools aims and goals.
14	SchoolMissionVisionNorms	Number	the school has common mission, vision and norms.
15	ParticipatoryPlans	Number	the school has participatory plans for school improvement and has identified the situations that need primary concern.
16	StudentParticipation	Number	Students' education participation is improved in the school.

3.2 Data Preparation for Analysis

Data preparation, is an intensive and time-consuming task necessary for successful data mining. As most real world datasets are not part of a data warehouse it most likely clcontain inconsistent,

incomplete and noisy data. Inconsistencies and incompletness in datasets Results the need for further data cleaning in order to make the data set ready for analysis.

3.2.1 Data Cleaning

Data cleaning has become an obligatory task to improve the quality of data so as to prevent garbage in garbage out and to improve the accuracy and efficiency of the data mining techniques.

The data cleaning tasks engage in the assortment of clean subsets of the data set to elevate the data quality to the point obligated by the chosen analysis techniques. Noisy values containing errors or outliers for attributes are also deleted and set to blank. For example “rural/urban” for school location attribute are cleaned. Incomplete recordes lacking attribute values are cleaned. Inconsistent values of records containing discrepancies are also detected and cleaned.

In order to handle missing values WEKA Replace Missing Values data filtering method is used. Since all the attributes are nominal their corresponding missing values are replaced by modal values.

3.2.2 Attribute Selection

Attribute selection is concerned with evaluating the relevance of attributes for analysis process by searching through all possible combinations of attributes in the data. Attributes having less significance for analysis purpose should be excluded to get effective analysis results. Therefore, From the 39 attributes of the original data set, only 27 attributes have been selected, which are thought to have significant contribution in appraising quality education factors on schools. Other attributes such as school name, ownership, location, zone, woreda, type and code are not selected as significant attributes for the analysis purpose as they only describe the schools. The selected 27 attributes are shown in table 2.

Table 2: list of selected attributes and their description

1	TeachingLearningFacilities	Number	The fulfillment of the necessary teaching learning facilities and buildings, supporting materials in the school.
2	FinancialEfficiencyforImpr	Number	The school’s financial efficiency for

	ovement		primary improvements in teaching learning process.
3	StaffEfficiency	Number	The school has efficient dean, teachers, and support staffs with appropriate education level and behavior.
4	SafeTeachingLearningEnviroment	Number	The school has provided safe teaching learning environment for the school society.
5	DevelopmentSociety	Number	the school has created education society that understands and is willing to practice the schools aims and goals.
6	SchoolMissionVisionNorms	Number	the school has common mission, vision and norms.
7	ParticipatoryPlans	Number	the school has participatory plans for school improvement and has identified the situations that need primary concern.
8	StudentParticipation	Number	Students' education participation is improved in the school.
9	StudentsEducationAcceptance	Number	Students showed improvement in their education acceptance, time management and creativity.
10	StudentsSentiment	Number	students are satisfied with the schools services, supported the school works, give respect for the school society, obey the schools rules and regulations.
11	TeachersEducationDeliveryPerformance	Number	Teachers deliver education in a way that is well planed, well supported with teaching

			aids and aiming to achieve high education results.
12	TeachersCapacity	Number	Teachers have enough knowledge and skill of the subject they teach and provide the subject in appropriate language and level that suits the students.
13	ModernTeachingMethods	Number	The school has created suitable conditions to provide modern and participatory way for teaching learning.
14	FemaleandSpecialNeed	Number	The school provides special support for female students and for those which has special needs that is needed to improve their education result.
15	CPD	Number	Teachers, deans and supervisors identified problems and prepared module to resolve the problems and participated on improvement programs.
16	DevelopmentWork	Number	School leaders, teachers, students, and support staffs are grouped in different perspectives to build education development society and participate in decision making processes and supervision.
17	CurriculumEvaluation	Number	Teachers evaluate the curriculum, its importance, participatory and it's consider ability of the students' level and need and give results and improvements.
18	StudentsEvaluation	Number	Students are correctly evaluated based on

			the curriculum and have been given appropriate results.
19	Plans	Number	The school leadership staffs controls plans to be executed on scheduled time and quality.
20	Standards	Number	The school has established a standard for using human, financial and material resources.
21	SchoolParentRelationship	Number	The school has strong relation with the students' parents and the society, motivates parents to actively participate.
22	DevelopmentGoals	Number	The school achieved educational participation and internal efficiency education development goals that are set nationally.
23	ResultImprovement	Number	Result in class, regional and national examinations has been improved with respect to the expected national and regional evaluation.
24	ResponsibilityandBehaviour	Number	It's practically proven that students has good behavior, with good cultural norms and are responsible on taking care of their environment.
25	SenseofResponsibility	Number	A good relationship and sense of responsibility is developed between the school teachers, leading staffs and supporting staffs.

26	ParentsRelationshipandSupport	Number	The school gains support from parents and the society by creating strong relationship with parents, the society and support organizations provide support for the school.
27	Grade	Number	The performance evaluation rank given for the school according to the performance result.

3.2.3 Data Set Format

Data provided for WEKA needs to be set in a file format that is acceptable for the software. Data records that are in comma separated format and Attribute Relation File Format (ARFF) are acknowledged by WEKA software.

Preparing the data in a format that can be recognized by the WEKA requires converting the dataset from Microsoft excel file format to a Comma Delimited (CSV) file by saving as a .csv file format. WEKA can process a data file either in .csv or .arff file format. For this study the data is provided to the WEKA in .csv format.

CHAPTER FOUR
4 DATA ANALYSIS AND EXPERIMENTATION

Data analysis is the process of evaluating data using analytical and logical reasoning to examine each component of the data provided. Data from various sources is gathered, reviewed, and then analyzed to form some sort of finding or conclusion.

In chapter three, the data is well understood, explored, selected and clean enough to be used for model building. This chapter presents the detailed data analysis carried out in selecting a modeling technique, implementation of the technique selected using the appropriate algorithms and evaluation of the models to select the best one for factors identification.

4.1 Modeling Technique

According to the CRISP data mining standard methodology, selecting the appropriate modeling technique is the first step in modeling. Decision tree, Neural Networks, Support Vector Machine, K-Means Clustering, Linear Discriminant Analysis and Linear Regression model; predictive modeling methods perform deduction on the current data by using some variables or fields in the data base to predict unknown or future values of other variables of interest.

WEKA data mining tool implements decision tree using AD Tree, Decision Stump, ID3 and J48 algorithms. Classifications rules are a popular alternative to decision trees in representing the structures that learning methods produce. The precondition of a rule is a series of tests just like the tests at nodes in decision trees. WEKA implements rule induction using Jrip, PART, OneR and ZeroR algorithms.

In this research work, J48 decision tree algorithm an easy to build and understand, which automatically handles interactions between variables and identifies important variables and the JRip rule induction algorithm which creates a series of if-then statements to utilizes a set of assertions and rules to act upon those assertions are applied for knowledge representation.

4.2 Model building

Model building involves the process of specifying model settings, Viewing model details, Testing the model and Evaluateing the model aiming to increase knowledge of the dataset. Followed by organizing and presenting the knowledge gained in a way that the customer can use it.

The cleaned and preprocessed amhara region inspection dataset is fed in to WEKA, having 8514 instances and 27 different attributes with four distinct values labeled as 1,2,3 and 4 as shown in figure 2.

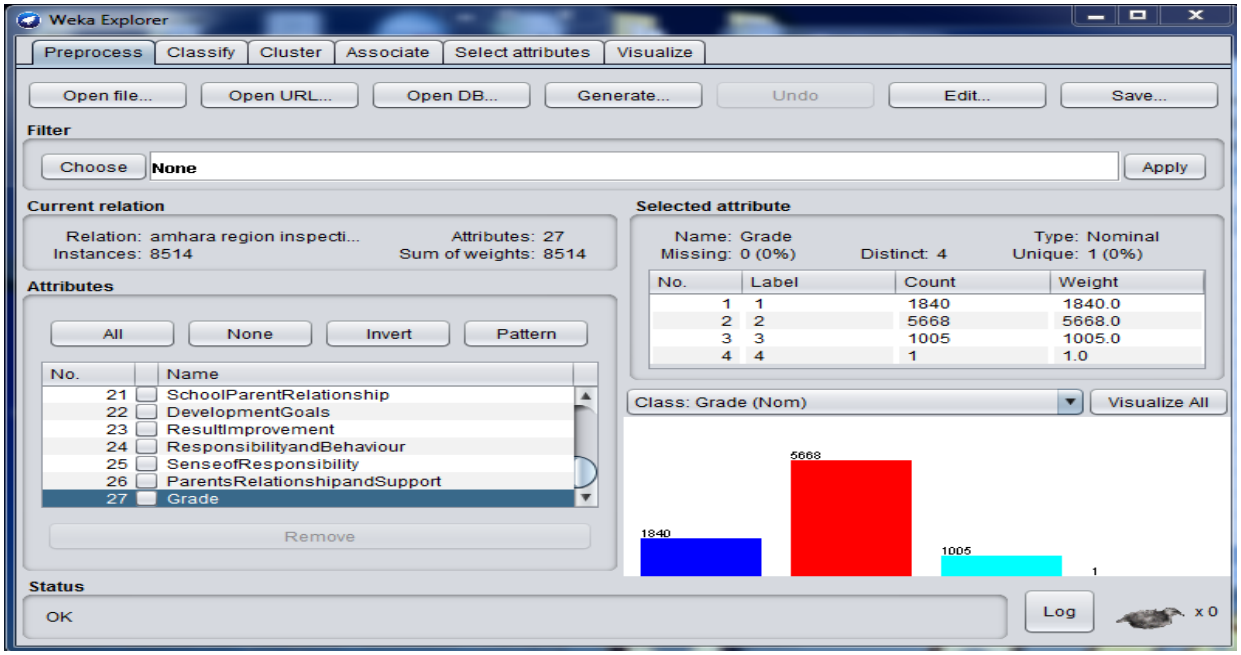


Figure 2: the dataset fed to WEKA for experiment

4.2.1 Experiment one

J48 algorithm

WEKA has implementation procedures of numerous classification and prediction algorithm to develop decision tree. J48 algorithm of decision tree technique is one of these algorithms which support both numeric and nominal predicators and nominal class attribute values.

J48 algorithm is an implementation of the C4.5 decision tree learner. The algorithm for induction of decision trees uses the greedy search technique to induce decision trees for classification. J48 has a facility of generating outputs both in tree form and rule sets. The set of rules are easier to understand as it shows the hierarchy of the determinant factors or attributes. J48 also accounts for missing values, decision trees pruning and derivation of rules. Thus j48 decision tree algorithm is selected among the decision tree techniques for this particular research. A brief explanation of how J48 algorithm works and description of parameters has been made in the previous chapter.

In this experiment the selected 27 attributes are used to build the model. Since the explorer generally chooses sensible defaults the J48 decision tree algorithm with all its default parameters is run on the dataset. The training and testing is done using tenfold cross validation.

The k-fold (k=10) cross validation test options is used; by doing so the partition and experiment could be more reliable. In this test option the accuracy estimate is the overall number of correct classifications from the k iteration divided by the total number of samples, which is k. After deciding the values of the parameters the algorithm is run to start building the model.

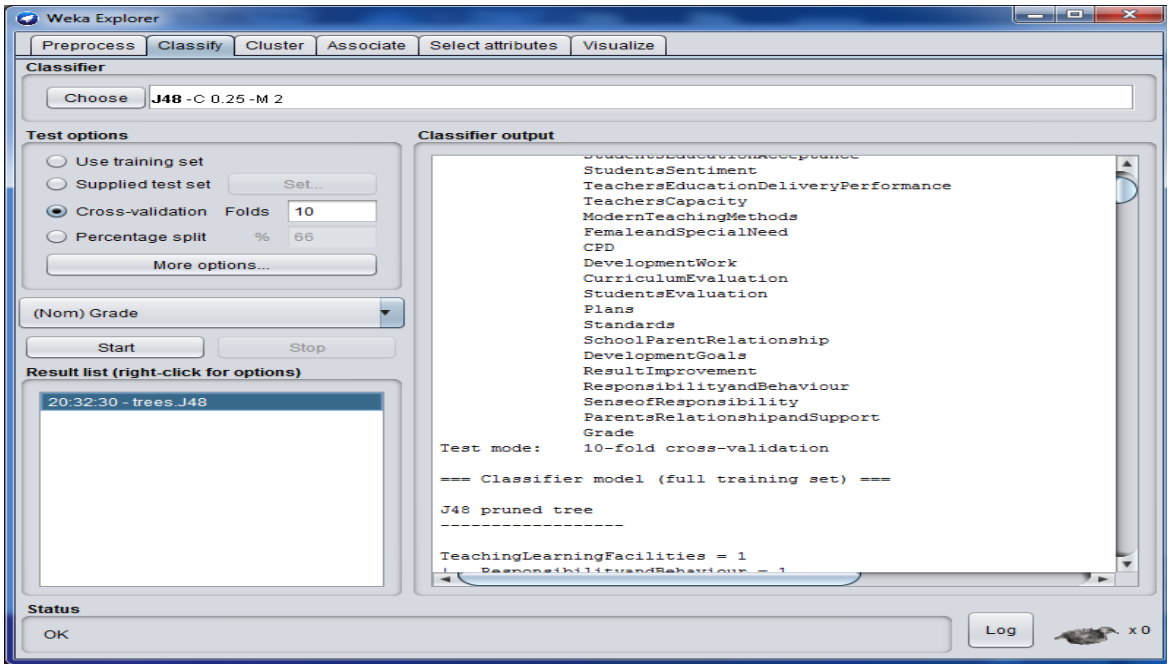


Figure 3:J48 classification algorithm out put in WEKA

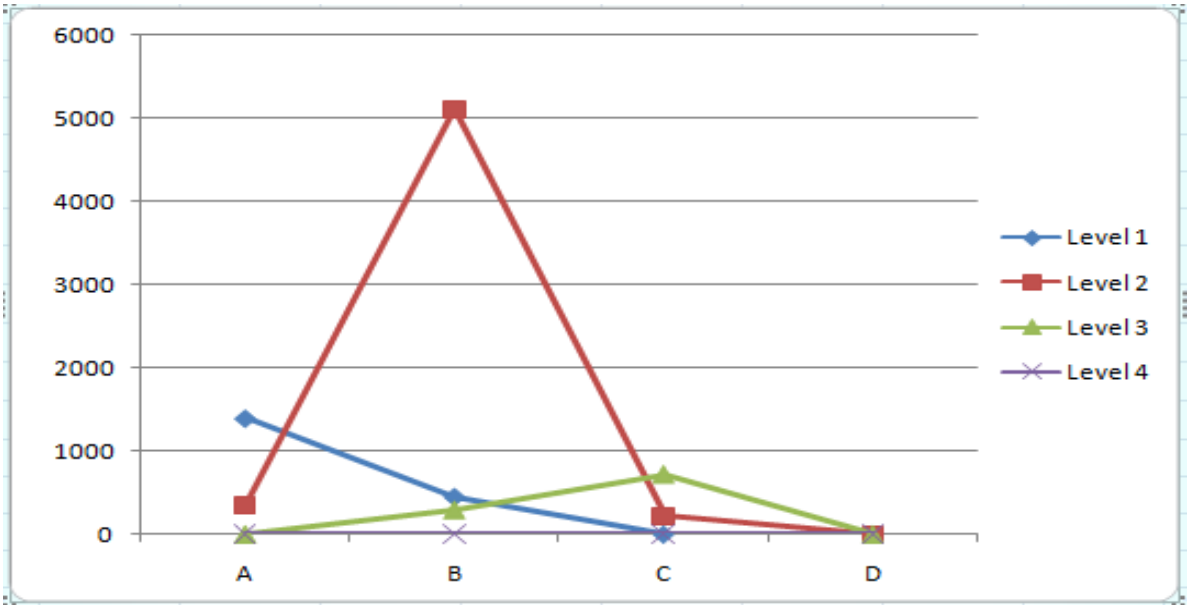


Figure 4: J48 confusion matrix summary

As indicated in figure 3, out of the 8,514 records of the dataset 7,209 records are correctly classified by the model and the model has an accuracy of 84.6723 %. The confusion matrix also shows that 1,389 instances out of 1,840 are correctly classified as 1 (poor performance), 5,106 instances out of 5,668 are correctly classified as 2 (on improvement) and 714 instances out of 1005 are correctly classified as 3 (good performance) and no instance is correctly classified as 4 (high performance) are correctly classified.

4.2.2 Experiment two

Attribute selection

In this experiment, WEKA's attribute selection feature is used for performing the J48 algorithm. Using the select attribute feature of WEKA and applying attribute selection Based on the ranking of attributes, all the 26 attributes are selected. As a result no need to run same experiment using J48 classification algorithm as it will give the same classification result as the experiment one. The following output is achieved from the attribute selection, which rank attributes based on their information gain. Based on the rank given by the algorithm: Responsibility and Behavior, Students Evaluation, Teaching Learning Facilities, Teachers Education Delivery Performance, Student

Participation are the top five attributes that are selected to be more determinant factors for quality education delivery of schools. Selected attributes: 24, 18, 1, 11, 8, 22, 25, 5, 19, 4, 21, 2, 13, 20, 9, 26, 16, 6, 7, 10, 17, 23, 14, 15, 12, 3: 26.

Table 3: WEKA Attribute Selection using Attribute ranking

Ranked	St.No.	Attribute Name
0.27	24	ResponsibilityandBehaviour
0.269	18	StudentsEvaluation
0.268	1	TeachingLearningFacilities
0.265	11	TeachersEducationDeliveryPerformance
0.264	8	StudentParticipation
0.26	22	DevelopmentGoals
0.246	25	SenseofResponsibility
0.245	5	DevelopmentSociety
0.245	19	Plans
0.242	4	SafeTeachingLearningEnviroment
0.239	21	SchoolParentRelationship
0.239	2	FinancialEfficiencyforImprovement
0.239	13	ModernTeachingMethods
0.238	20	Standards
0.235	9	StudentsEducationAcceptance
0.227	26	ParentsRelationshipandSupport

0.224	16	DevelopmentWork
0.211	6	SchoolMissionVisionNorms
0.209	7	ParticipatoryPlans
0.208	10	StudentsSentiment
0.205	17	CurriculumEvaluation
0.183	23	ResultImprovement
0.143	14	FemaleandSpecialNeed
0.142	15	CPD
0.127	12	TeachersCapacity
0.107	3	StaffEfficiency

4.2.3 Experiment Three

JRip algorithm

JRip rule induction algorithm is run on education dataset with tenfold cross validation to come up with significant rules that can be used for defining quality education factors.

Domain experts are consulted intensively in evaluating the significance of the rules. As a result the rules generated based on the twenty seven attributes. The performance of the algorithm in generating the rules is 84.8015 %of accuracy. Figure 4 shows some of the rules generated by the JRip algorithm. The rest of the rules can be found in Appendix II.

Using JRip algorithm, the experiment showed that Teaching Learning Facilities, Financial Efficiency for Improvement, Student Participation, Teachers Education Delivery Performance, Students Sentiment,Responsibility and Behavior are the most important variables to classify records to their predefined class.

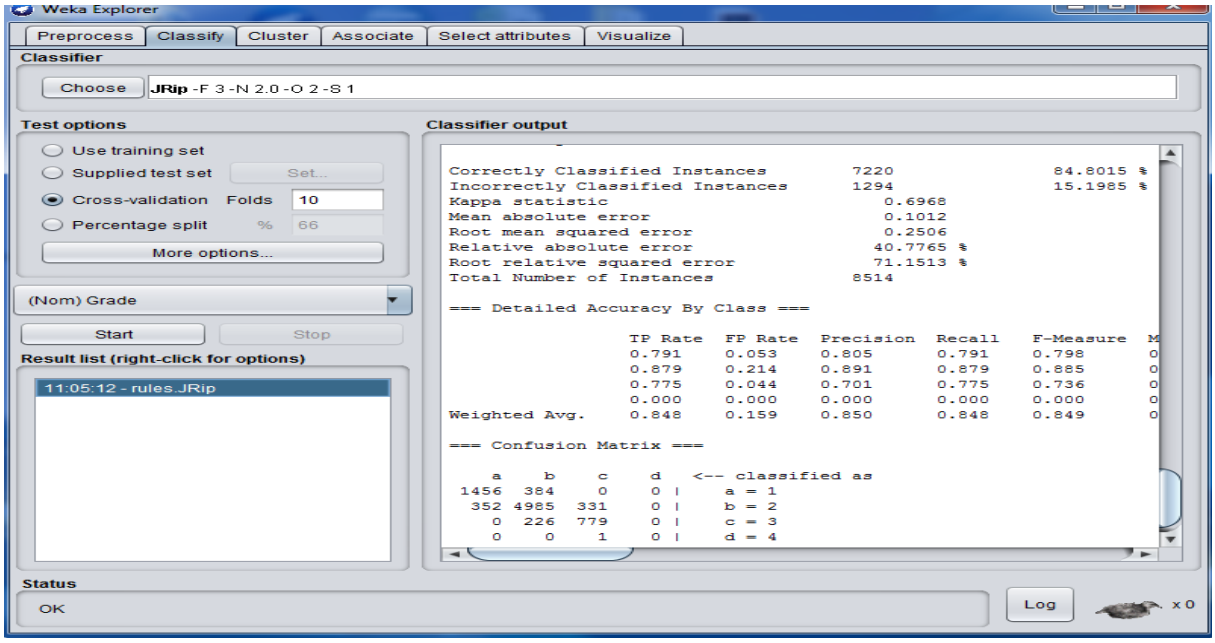


Figure 5: jRip classification algorithm output (partial view)

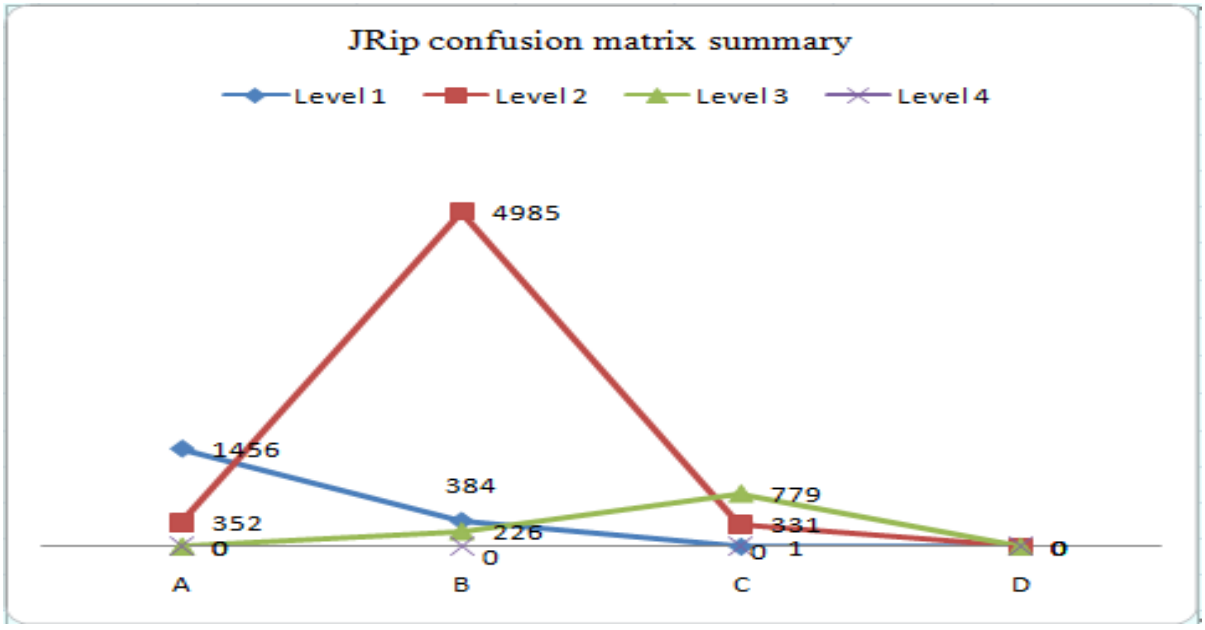


Figure 6: JRip algorithm classification confusion matrix

As indicated in figure 5, out of the 8,514 records of the dataset 7,220 records are correctly classified and the model has an accuracy of 84.8015 %. The confusion matrix also shows that 1,456 out of 1,840 are classified as 1 (poor performance), 4985 out of 5,668 classified as 2 (on improvement) and 779 out of 1005 are classified as 3 (good performance) and no 4 (high performance) are correctly classified.

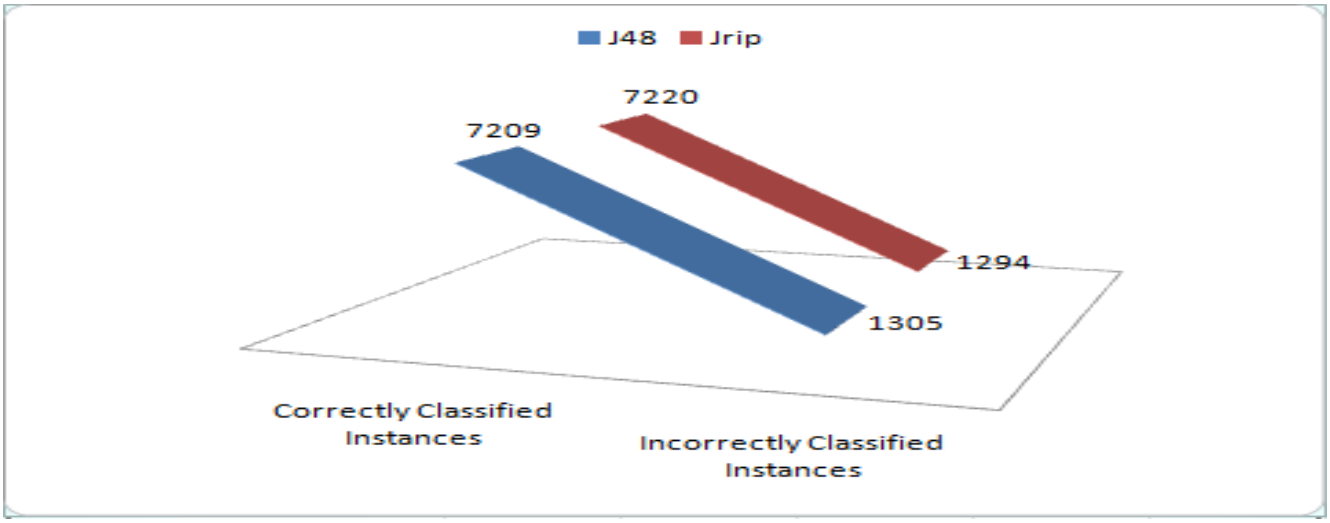


Figure 7: correctly and incorrectly classified instances of J48 and JRip algorithms

4.3 Models Evaluation

Evaluation is the process of measuring the performance, which is important for the prediction model how it works in the future. It is an integral part of many learning methods, to find the model that best represents the expected outputs. For such expected Finding of accuracy by itself doesn't tell everything about the efficiency of predicted model. Therefore, WEKA's experiment has been utilized to implement j48 decision tree learners and the JRip rule induction algorithm to analyze the models performance. The experiment type is a ten-fold cross-validation and model parameters have been used.

Both algorithms are executed at the same time, on the same data set, and equal number of attributes (27 attributes). After processing the necessary setup the algorithms are used for the experiment. To analyze

the finding the percent-correct comparison measurement factor is selected from the comparison field box. and the two models have been analyzed by the testing data automatically. The comparison result showed that the JRip algorithm performs better than the J48 algorithm with an accuracy of 84.80 % and 84.67% respectively. As it is shown on figure below, the result of WEKA's environment window of the selected algorithms (J48 and JRip).

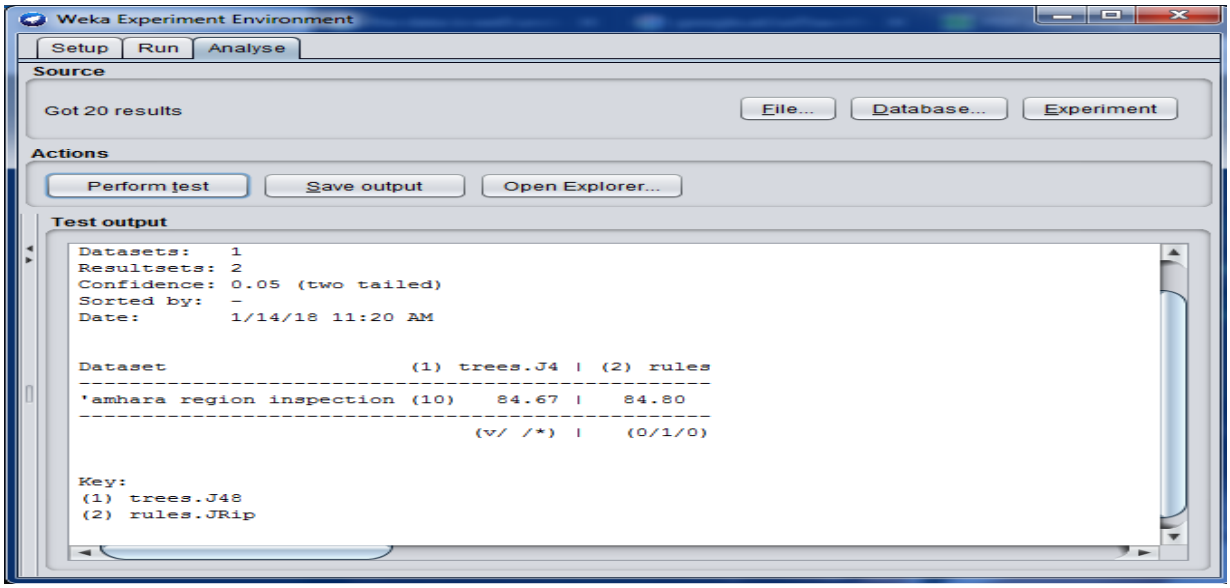


Figure 8: Evaluation of the classification models

4.4 Discussion of Results

Discussion of results presents the out puts of the algorithms after processing the data and an interpretation of the results. Some of the works which are done using data mining techniques are presented along with the description of the methods they use and the perfomance of the algorithms they used.

Table 4: Descusion of other research results

Title	Authors	Dataset	Framework	Algorithm	Outperforming algorithm
The Prediction of Students' Academic Performance Using Classification Data Mining Techniques	Fadhilah Ahmad, Nur Hafieza Ismail and Azwa Abdul Aziz	students' demographics, previous academic records, and family background information collected from 8 year period	framework for predicting students' academic performance of first year bachelor students	Decision Tree, Naïve Bayes, and Rule Based	Rule Based71.3%
Prediction Of Student Performance Using Weka Tool	Gurmeet Kaur, Williamjit Singh	52 data records	Prediction of student Performance	Naïve bayes and J48 decision tree classification techniques	Naïve bayes 63.59 % and j48 61.53%
Developing prediction model of loan Risk in banks using data mining	Aboobyda Jafar Hamid and Tarig Mohammed Ahmed	A collection of data from banking sector	a new model for classifying loan risk in banking sector	j48, bayesNet and naiveBayes	j48 78.3784 % bayesNet 77.4775 % naiveBayes 73.8739 %
Student performance prediction using classification data mining techniques	Muhammed Salman Shamsi and Jhansi Lakshmi	students data , academics, socio-economic state, psychological state	predict the performance of students in terms of grades and dropout for a subject	Naïve Bayes, LibSVM, J48, random forest, and Jrip	Jrip

As we can see from the above table, the works done by other researchers using data mining techniques on WEKA enviroment have different algorithm preferences depending on the objective of the research and the dataset used. Different algorithms outperform in the expermentation for each research work. This shows us that the performances of the algorithms varry depending on the data size and the parameters that are used for the research work.

The result of our study by applying decision tree J48 algorithm and rule based jRip algorithm is presented. The findings are explained and presented in detail.

J48 Algorithm

In the output of the J48 algorithm, the summary of the dataset, and the default 10 fold cross-validation used in which it is used to evaluate the quality education performance; then comes a pruned decision tree in textual form. The first split is on the "Teaching Learning Facilities" attribute, and then, at the second level, the split is on "Responsibility and Behavior" and it goes on splitting on the most important attributes based on the measure of purity the algorithm uses until there is no more attribute left to split on or the data is completely classified. In the tree structure, a column introduces the class label that has been assigned to a particular leaf, followed by the number of instances that reach to the subsequent leaf, expressed as a decimal number. Because of the way, the algorithm uses fractional instances to handle missing values. For example, an expression (543.0/21.0) from the fourth leaf node in Appendix I imply a total of 543 instances reached that leaf node out of which 21 are classified incorrectly.

The next part of the output gives estimates of the tree's predictive performance. In this case attributes are obtained using stratified cross validation with 10 folds. It can also be seen that out of the 8,514 instances, 7,209 (84.6723 %) are correctly classified and 1,305 (15.3277 %) are incorrectly classified in the cross-validation.

In addition to the classification error, the evaluation module also outputs the Kappa statistic, the mean absolute error, and the root mean-squared error of the class probability estimates assigned by the tree. The root mean squared error is the square root of the average quadratic loss. The mean absolute error is calculated in a similar way using the absolute instead of the squared difference. Finally, the confusion matrix at the bottom of the output shows that 1,389 out of 1,840 1(poor

performance), 5,106 out of 5,668 2(improving), 714 out of 1,005 3(good performance) are correctly classified and no 4(high performance) is classified correctly.

jRip Algorithm

The classifier used the 27 attributes to construct rules and provide the class predicted by the model. The numeric values which appeared in bracket next to the class label indicates the number of correctly and incorrectly classified records respectively. For example Rule 1

(Teaching Learning Facilities = 3) and (Financial Efficiency for Improvement = 3) and (Student Participation = 3) and (Teachers Education Delivery Performance = 3) and (Students Sentiment = 3) and (Responsibility and Behavior = 3) => Grade=3 (85.0/2.0)

This can be interpreted as there are 85 records in the dataset that are having Grade 3 when Teaching Learning Facilities = 3 and Financial Efficiency for Improvement = 3 and Student Participation = 3 and Teachers Education Delivery Performance = 3 and Students Sentiment = 3 and Responsibility and Behavior = 3; which shows that having a good result in these attributes will result having a good Grade. 2 records are misclassified to this rule.

As the result indicates Teaching Learning Facilities, Financial Efficiency for Improvement, Student Participation, Teachers Education Delivery Performance, Students Sentiment, Responsibility and Behavior Are the most determinant parameters that need to be focused so the schools and quality education experts have to focus on these parameters by collaboration with the regional education bureau.

CHAPTER FIVE
5 CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Organizations and citizens are trying to assure “education for everyone”. However since it is a fact that not only “education for everyone” but also quality education is the key to development; quality education should be delivered for everyone to make students and graduates effective on their field.

Anchored in the hard effort of joint researchers, today’s science is dependent to a great extent on data analysis. As the outlay for processing power and storage is diminishing data storage happens to be undemanding and economical. Accordingly the amount of data accumulated in databases is escalating rapidly. Different Organizations accumulate enormous amount of data for years. However, vast amount of these data sources create unfeasibility for human experts to come up with interesting information or patterns that will help in the practical decision making process since experts do not have time to go through all records and data collections in their organizations. Moreover, the amount of data is bulky having several variables; it is extremely difficult to visualize patterns and relationships. Human Experts need filtered and simplified data from their large amount of records. Knowledge discovery systems could help experts to pass correct decisions on their daily activity or improve their future plan. Data mining technology; a Data analytics tool with capability of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making finds out hidden pattern from vast amount of data.

In this work, we have proposed to identify the determinant factors for enhancing and defining quality education performance using data mining techniques. We apply the decision tree and Rule induction predictive data mining techniques to identify the most determinant factors for education quality from the education data set. To achieve this goal: the CRISP-DM 1.0 standard data mining methodology has been adopted. The Experiments are run on WEKA enviroment using J48 decision tree algorithm and JRip rule induction algorithms to build models which identify the most determinant factors in quality education problem of the schools.

The classification algorithms J48 and Jrip produces different rules that provides the most determinant factors from the given dataset, with an accuracy of 84.67 and 84.80 respectively.

The models comparison performed using WEKA's experimenter to select the best classification model for the dataset showed that JRip rule induction algorithm with 84.80 % accuracy outperforms J48 decision tree algorithm with an accuracy of 84.67 %. Based on the comparison result the rules obtained from JRip rule induction algorithm are taken as the determinant factors for education quality. this includes; Teaching Learning Facilities, Financial Efficiency for Improvement, Student Participation, Teachers Education Delivery Performance, Students Sentiment, Responsibility and Behavior.

In general, encouraging results are obtained by employing both decision tree and rule induction technique, and the rules generated by J48 and JRip algorithm are easily understandable by subject experts in the department. Thus, the results obtained in this research have proved the applicability of data mining in defining and enhancing education quality. More specifically it provides valuable and help in developing new methods to increase quality education, particularly in choosing the most determinant factors for quality education.

However, the research scope did not allow making it wide to include even some other external factors that can enable to have a quality data which in turn increases the accuracy of the models to be built.

5.2 Recommendation

This research work is conducted mainly for academic purpose. In this research, we focused on the general parameters of education quality performance factors.

However, Other parameters which are specific to the ownership(public, private) and location(rural, urban) of the schools can be considered to define quality education factors.

REFERENCES

Aboobyda Jafar Hamid, T. M. (March 2016). Developing Prediction Model of Loan Risk In Banks Using Data Mining. *An International Journal (MLAIJ) Vol.3, No.1, .*

Ahmad Fadhilah Aziz, N. H. (2015). *The Prediction of Students' Academic Performance Using Classification Data Mining Techniques.* Malaysia: Universiti Sultan Zainal Abidin (UniSZA).

CRISP-DM 1.0. (August 2000).

Dr.Kishori. (June 2014). *Comparative Study of Data Mining Tools; Volume 4, Issue 6,.* Kalpana Rangra: International Journal of Advanced Research in Computer Science and Software Engineering.

FangWeiping, W. Y. (2013). The Development of Data Mining. *International Journal of Business and Social Science Vol. 4 No. 16; December 2013 .*

Fayyad, P.-S. S. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *AAAI (www.aaai.org) .*

goopta, c. (Oct 2014). *Six of the Best Open Source Data Mining Tools.*

Gurmeet Kaur, W. S. (January 2016). Prediction Of Student Performance Using Weka Tool. *An International Journal of Engineering Sciences, Vol. 17.*

Himani S., Sunil K. (April 2016). A Survey on Decision Tree Algorithms of Classification in Data Mining. *An International Journal of Science and Research,* Vol 5.

Jiawei Han, M. K. (2006). Data Mining Concepts and Techniques.

MOE. (2014). Educational Inspection Framework.

Mohammed I Al-Twijri, A. Y. (2015). *A New Data Mining Model Adopted for Higher Institutions.* International Conference on Communication, Management and Information Technology (ICCMIT).

Nissan Levin, J. (1999). Predictive modeling using segmentation. *Journal of Interactive Marketing March 2001 .*

Oprea, C. (2014). Performance Evaluation of The Data Mining Classification. *Academica Brancusi Publisher, Issn 2344 – 3685/Issn-L 1844 - 7007 .*

Pooja Thakar, A. M. (2015). Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue. *International Journal of Computer Applications © 2015 by IJCA Journal Volume 110 – No.15 .*

Qasem A. Al-Radaideh, A. A.-S. (August 2011). A classification model for predicting the suitable study track for school students. *IJRRAS* .

Ryan S.J.d.Baker. (2010). Data mining for education. *International encyclopedia of education 7 (3), 112-118* .

S.Visalaxi, S. S. (April 2015). Efficient classifier for predicting Students knowledge level using data Mining techniques . *International Journal of Advance Research In Science And Engineering IJARSE, Vol. No.4* .

Saini, P. (n.d.). Building a classification model for enrollment in higher educational courses using data mining techniques.

S. Sumathi, S.N. Sivanandam. (2006). Introduction to Data mining and Its Application. *Springer* .

Sumit Garg, A. k. (July 2013). Comparative Analysis of Data Mining Techniques on Educational Dataset. *International Journal of Computer Applications (0975 – 8887) Volume 74– No.5* .

Umesh Kumar Pandey S. Pal. (2011). Data Mining: A prediction of performer or underperformer using classification. *International Journal of Computer Science and Information Technologies, Vol. 2 (2)* .

UNICEF. (June 2000). Defining Quality in Education.

APPENDICES

Appendices 1

J48 algorithm output (sample)

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: amhararegioninspection weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last

Instances: 8514

Attributes: 27

TeachingLearningFacilities
FinancialEfficiencyforImprovement
StaffEfficiency

SafeTeachingLearningEnviroment

DevelopmentSociety

SchoolMissionVisionNorms

ParticipatoryPlans

StudentParticipation

StudentsEducationAcceptance

StudentsSentiment

TeachersEducationDeliveryPerformance

TeachersCapacity

ModernTeachingMethods

FemaleandSpecialNeed

CPD

DevelopmentWork

CurriculumEvaluation

StudentsEvaluation

Plans

Standards

SchoolParentRelationship

DevelopmentGoals

ResultImprovement

ResponsibilityandBehaviour

SenseofResponsibility

ParentsRelationshipandSupport

Grade

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

TeachingLearningFacilities = 1

| ResponsibilityandBehaviour = 1

| | Plans = 1

| | | ResultImprovement = 1: 1 (543.0/21.0)

| | | ResultImprovement = 2

| | | | SchoolParentRelationship = 1

| | | | | ParticipatoryPlans = 1: 1 (53.0)

| | | | | ParticipatoryPlans = 2

| | | | | | CurriculumEvaluation = 1: 1 (20.0)

| | | | | | CurriculumEvaluation = 2

| | | | | | | FinancialEfficiencyforImprovement = 1: 1 (20.0/2.0)

| | | | | | | FinancialEfficiencyforImprovement = 2

| | | | | | | | StudentParticipation = 1: 1 (3.0/1.0)

| | | | | | | | StudentParticipation = 2: 2 (5.0)

| | | | | | | | StudentParticipation = 3: 2 (0.0)

| | | | | | | | StudentParticipation = 4: 2 (0.0)

| | | | | | | FinancialEfficiencyforImprovement = 3: 1 (0.0)

| | | | | | | FinancialEfficiencyforImprovement = 4: 1 (0.0)

| | | | | CurriculumEvaluation = 3: 2 (1.0)
| | | | | CurriculumEvaluation = 4: 1 (0.0)
| | | | ParticipatoryPlans = 3
| | | | | StudentsEducationAcceptance = 1: 1 (4.0/1.0)
| | | | | StudentsEducationAcceptance = 2: 2 (2.0)
| | | | | StudentsEducationAcceptance = 3: 1 (0.0)
| | | | | StudentsEducationAcceptance = 4: 1 (0.0)
| | | | ParticipatoryPlans = 4: 1 (2.0)

Number of Leaves: 697

Size of the tree: 929

Time taken to build model: 0.4 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	7209	84.6723 %
Incorrectly Classified Instances	1305	15.3277 %

=== Confusion Matrix ===

a	b	c	d	<-- classified as
1389	450	1	0 |	a = 1
346	5106	216	0 |	b = 2
0	291	714	0 |	c = 3
0	0	1	0 |	d = 4

Appendices 2

JRip algorithm output (sample)

=== Run information ===

Scheme: weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1

Relation: amhara region inspection-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last

Instances: 8514

Attributes: 27

TeachingLearningFacilities

FinancialEfficiencyforImprovement

StaffEfficiency

SafeTeachingLearningEnviroment

DevelopmentSociety

SchoolMissionVisionNorms

ParticipatoryPlans

StudentParticipation

StudentsEducationAcceptance

StudentsSentiment

TeachersEducationDeliveryPerformance

TeachersCapacity

ModernTeachingMethods

FemaleandSpecialNeed

CPD

DevelopmentWork

CurriculumEvaluation

StudentsEvaluation

Plans

Standards

SchoolParentRelationship

DevelopmentGoals

ResultImprovement

ResponsibilityandBehaviour

SenseofResponsibility

ParentsRelationshipandSupport

Grade

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

JRIP rules:

==========

(TeachingLearningFacilities = 3) and (FinancialEfficiencyforImprovement = 3) and (StudentParticipation = 3) and (TeachersEducationDeliveryPerformance = 3) and (StudentsSentiment = 3) and (ResponsibilityandBehaviour = 3) => Grade=3 (85.0/2.0)

(DevelopmentGoals = 3) and (TeachersEducationDeliveryPerformance = 3) and (ModernTeachingMethods = 3) and (SchoolParentRelationship = 3) and (StudentsEducationAcceptance = 3) and (ParentsRelationshipandSupport = 3) and (SenseofResponsibility = 3) => Grade=3 (39.0/1.0)

(TeachingLearningFacilities = 3) and (ModernTeachingMethods = 3) and (SenseofResponsibility = 3) and (FemaleandSpecialNeed = 3) => Grade=3 (44.0/0.0)

Number of Rules: 89

Time taken to build model: 19.4 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	7220	84.8015 %
Incorrectly Classified Instances	1294	15.1985 %

=== Confusion Matrix ===

a	b	c	d	<-- classified as
1456	384	0	0	a = 1
352	4985	331	0	b = 2
0	226	779	0	c = 3
0	0	1	0	d = 4

Appendices 3

Schools evaluation criteria

I. Input

i. School facility, physical, human and financial resources.

1 the school has fulfilled the necessary teaching learning facilities and buildings, supporting materials

INDICATORS:

1.1 schools learning and other service providing buildings are built according to the standard and considering those who have special need.

1.2 The school has provided students book, teacher's books and other supporting books, brail and other materials.

1.3 The school has provided libraries, laboratories and sport field.

1.4 education training policy, national and regional programs, other related schools policies are fulfilled in the school.

2 the school is financially efficient for primary improvements in teaching learning process.

INDICATORS:

2.1 The school has accepted the block grant and used it effectively according to the standard.

2.2 The school has accepted the school grant and used it effectively according to the standard.

2.3 The school has gained benefits in money, man power, and item from student's parents and the society.

2.4 The school has developed its financial income by creating internal income sources.

2.5 The school has gained income from NGO's and other individuals (former students, the society etc.)

2.6 The school has well organized finance documents.

3 the school has efficient dean, teachers, and support staffs according to the standard.

INDICATORS:

3.1 The school dean and teachers has work license and appropriate education certificate.

3.2 The school has support staffs which has appropriate education certificates and which are appropriate for the level.

3.3 The school has staff which gives guidance and counseling.

3.4 The school has staffs which are trained for special needs.

ii. Suitable environment for education

4 the school has provided safe teaching learning environment for the school society.

INDICATORS:

4.1 The school has sufficient area and width according to the standard.

4.2 The school has legal documents about its placement and area.

4.3 The school buildings are suitable for teaching and learning and are built by considering those who has special need.

4.4 The school compound has boundary.

4.5 The school is free from disturbing conditions which affect teaching learning process.

4.6 The school has toilets for teachers, students and employees which are enough, standardized, cleaned every time, with water and soap supplies and separated according to gender.

4.7 The school has clean water for drink.

5 the school has developed well organized education development society.

INDICATORS:

5.1 situations are created that allows practicing the schools aim and mission.

5.2 The school has created education society that understands and willing to practice the schools aims and goals.

5.3 The school has created situations that allow effective jobs to be done.

6 the school has common mission, vision and norms.

INDICATORS:

6.1 The schools leadership allows participation of concerned parties and created aim, mission and norms.

7 the school has participatory plans for school improvement

INDICATORS:

7.1 The school has identified the situations that need primary concern.

7.2 The school has prepared 3 years strategic plan with concerned parties.

iii. Leadership

II. Process

i. Teaching and Learning Learning

8 students learning participation is improved.

INDICATORS:

8.1 students do their assignments conservatively.

8.2 students actively participate by asking and answering questions.

8.3 students are working in network and help each other.

8.4 students are engaged in clubs and are actively participating.

8.5 students are participating in students' counseling.

9 students showed improvement in their education acceptance.

INDICATORS:

9.1 students use their time effectively.

9.2 students are creative and are solving problems in their own initiative.

9.3 students give equal value for all subjects.

9.4 students are well aware that cheating is forbidden.

10 students has good sentiment for their school.

INDICATORS:

10.1students are satisfied with the schools services.

10.2 students have supported the school works.

10.3 students have evaluated their teachers appropriately.

10.4 students give respect for the school society.

10.5 students obey the schools rules and regulations.

Teaching

11 teachers deliver education in a way that is well planned, well supported with teaching aids and aiming to achieve high education results. INDICATORS:

11.1 teachers course guidebook contains the aim, content of the subject and method of teaching.

11.2 teachers have prepared teaching learning materials and have been using them.

11.3 teachers have used modern technologies for teaching learning process (radio, TV, plasma, computer).

11.4 teachers have used laboratories for the subject they teach.

11.5 teachers encourage their students to use local materials and innovate new things to make them fruitful in science and technology.

11.6 teachers give tutorials for their students in order to improve their result.

12 teachers have efficient knowledge of the contents of the subject they teach.

INDICATORS:

12.1 teachers have enough knowledge and skill of the subject they teach.

12.2 teachers provide the subject in appropriate language and level that suits the students.

12.3 teachers provide the core ideas of the subject in a clear and easy way.

13 the application of modern and suitable teaching methods by teachers improve students' educational participation.

13.1 teachers use different participatory methodologies to make students researchers, innovators and problem solvers.

13.2 The school leaders have created suitable conditions to provide modern and participatory way for teaching learning.

13.3 teachers make students to learn individually and in group depending on its appropriateness.

13.4 teachers provide special support for female students.

13.5 teachers provide special support for those who need special treatment.

13.6 teachers have done practical researches to solve the teaching learning problem.

14 the school provides special support for female students and for those which has special needs.

14.1 The school has recorded data about students who need special treatment.

14.2 The school provided special support to improve the result of students which need special treatment.

14.3 The school provided special support to improve female students result.

15 teachers and supervisors performed continuous CPD.

INDICATORS:

15.1 teachers, deans and supervisors identified problems and prepared module to resolve the problems and participated on improvement programs at least 60 hours a year.

15.2 advising teachers are assigned for new teachers to give them induction. 16 teachers and supporting staffs are working together for development. INDICATORS:

16.1 The school leaders, teachers, students, and support staffs are grouped in different perspectives to build education development society and participate in decision making processes and supervision.

16.2 The school leaders, teachers and support staffs have good behavior, have respect for their job, and are willing to support the school.

ii. Curriculum

17 teachers evaluate the curriculum if it is valuable, participatory and if it considers the students level and need, gives results and improve. INDICATORS:

17.1 teachers know the current working education curriculum.

17.2 The subjects that the teachers give are prepared in the national and regional curriculum.

17.3 The subjects and curriculum materials are evaluated with respect to the students' development level and interest.

iii. Assessment

18 the students are correctly evaluated and have been given appropriate results.

INDICATORS:

18.1 The schools evaluation is prepared based on the curriculum and is prepared in table of specifications.

18.2 students are evaluated by exams prepared in regional/city administrations, zone/woreda.

18.3 teachers use continuous assessment to evaluate students result.

18.4 teachers give support by reviewing students result.

18.5 teachers give students evaluation results and give support for students to improve their result.

18.6 The school acknowledges students result to their parents and accept ideas.

iv. Leadership and management

19 the school leadership staffs controls plans to be executed on scheduled time and quality.

INDICATORS:

19.1 The school makes sure that the society has created well organized development plans and gives solution for the problems.

19.2 The education improvement committee organized in the school controls the school improvement program and gives support.

19.3 The continuous assessment committee controls trainings on continuous assessment improvement and identifies and gives support on what needs improvement.

19.4 The school leadership controls and supports the teaching learning process and clubs plan

19.5 The school appreciates and acknowledges bodies which work better.

20 the school has established a standard for using human, financial and material resources.

INDICATORS:

20.1 The school has organized

20.2 teachers are assigned to teach on the subjects they are trained.

20.3 deans and support staffs are assigned on the job they are trained.

20.4 The school buildings, facilities and additional materials are appropriately used.

20.5 The school budget is properly used for primary education improvement plans.

v. Engagement with parents and the community

21 the school has strong relation with the students parents and the society.

INDICATORS:

21.1 The school appreciates parents to actively participate on the school and class levels.

21.2 The school gives information continuously for parents and the society about students' education acceptance, result, behaviors, finance usage and other issues.

21.3 parents help students to work on their study at home.

21.4 parents actively participate on different groups in the school.

21.5 The school is beneficiary for the society.

21.6 different indicators show that parents are satisfied by the schools work.

III. Output

i. school and students attainment

22 the school achieved educational participation and internal efficiency education development goals that are set nationally.

INDICATORS:

22.1 children whose age is appropriate for education are attending schools.

22.2 The school has met its packaged participation plan.

22.4 The school has met its student gender limitation plans.

22.5 The schools dropout rate has decreased according to the plan.

22.6 The schools under level achievement of students have decreased according to the plan.

23 students result in class, regional and national examinations has been improved with respect to the expected national and regional evaluation. INDICATORS:

23.1 All students achieved in every subject 50% and above.

23.2 As the result of the schools special support for female students they are achieving 50% and above in every subject they take.

23.3 As the result of the schools special support for those who need special treatment, they are achieving 50% and above in every subject they take.

23.4 students' national and regional exam results are satisfactory according to the schools plan.

ii. students personal development

24 it's practically proven that students has good behavior, with good cultural norms and are responsible on taking care of their environment.

INDICATORS:

24.1 Students have good behavior, respects the school society.

24.2 students take care of the school properties.

24.3 students know and practice the schools norms, rules and regulations.

24.4 students accept their differences and solve their differences by negotiation.

24.5 students take care of their school and environment.

25 a good relationship and sense of responsibility is developed between the school teachers, leading staffs and supporting staffs.

INDICATORS:

25.1 teachers and support staffs respect for students improves students education interest.

25.2 there is a good working environment and relationship between schools teachers, leadership and support staffs.

25.3 The schools teachers, leadership and support staffs are responsible for their action.

iii. parents and society participation

26 creating strong relationship of the school with parents, the society and support organizations provide support for the school.

INDICATORS:

26.1 The school gains support from parents and the society.

26.2 parents and societies participation improvement makes the school perform better.