Prediction Model for School Readiness

Iyad Suleiman[1], Maha Arslan, Reda Alhajj, Mick Ridley

Abstract— Study the school readiness is an interesting domain that has attracted the attention of the public and private sectors in education. Researchers have developed some techniques for assessing the readiness of preschool kids to start school. Here we benefit from an integrated approach which combines data mining and social network analysis towards a robust solution.
The main objective of this study is to explore the socio-demographic variables (age, gender, parents' education, parents' work status, and class and neighbourhood peers influence) and achievement data (Arithmetic Readiness, Cognitive Development, Language Development, Phonological Awareness), data that may impact the school readiness.
This paper proposes to apply data mining techniques to predict school readiness. Real data on 306 preschool children was used from 4 different elementary schools: (1) Life school for Creativity and Excellence a private school located in Ramah village, (2) Sisters of Saint Joseph missionary school located in Nazareth, (3) Franciscan missionary school located in Nazareth and (4) Al-Razi public school located in Nazareth, and white-box classification methods, such as *induction rules* were employed. Experiments attempt to improve their accuracy for predicting which children might fail or dropout by first, using all the available attributes; next, selecting the best attributes; and finally, rebalancing data and using *cost sensitive classification*. The outcomes have been compared and the models with the best results are shown.

Index Terms— Data Mining, School Readiness, WEKA, . Life school for Creativity and Excellence, Machine Learning, Social Network Analysis, Prediction, Data Science.

1 Introduction

PARENTS and policy makers believe that children who start kindergarten with stronger cognitive and social skills are more likely to succeed in school. Research indicates that children enter school with a range of skills across five essential school readiness domains (i.e., language, cognition, social-emotional, approaches to learning, and health), but extant research has not systematically examined how skills in part or all five areas combine to predict school outcomes. Some evidence suggests that skills within in a domain (e.g., math) tend to be good predictors of the continued acquisition of those skills [21]. In addition, there is some evidence that skills in one area are important for later school outcomes in another area. For example, it is widely believed that children with stronger attention and social skills at school entry show faster acquisition of academic skills because they can sit and listen in the classroom [10]. In addition to the question of whether it is possible to have significant prediction across as well as within developmental domains from school entry through later schooling, another question is whether there are a set of skills at school entry that allow more disadvantaged children to catch up with more advantaged peers. There is growing interest in these questions at all levels, as educators and policymakers try to address how to support children's school success and monitor their overall development in a meaningful way. Currently, there is little research that examines trajectories of growth within and across multiple school readiness domains.

2 Educational Data Mining

Educational Data Mining (EDM) is the application of Data Mining (DM) techniques to educational data, and so, its objective is to analyze these types of data in order to resolve educational research issues [6].
DM can be defined as the process involved in extracting interesting, interpretable, useful and novel information from data [27].
It has been used for many years by businesses, scientists and governments to sift through volumes of data like airline passenger records, census data and the supermarket scanner data that produces market research reports [34].
Educational Data Mining (EDM) is an emerging multidisciplinary research area, in which methods and techniques for exploring data originating from various educational information systems have been developed. EDM is both a learning science, as well as a rich application area for data mining, due to the growing availability of educational data. EDM contributes to the study of how students learn, and the set- tings in which they learn. It enables data-driven decision making for improving the current educational practice and learning material [16].
On one hand, the increase in both instrumental educational software as well as state databases of student information has created large repositories of data reflecting how students learn [38]. On the other hand, the use of the Internet in education has created a new context known as e-learning or web-based education in which large amounts of information about teaching-learning interaction are endlessly generated and ubiquitously available [17]. All this information provides a gold mine of educational data [53]. EDM seeks to use these data repositories to better understand learners and learning, and to develop computational approaches that combine data and

- [1] *Iyad Suleiman, Ph.D. is currently a lecturer in Shenkar College of Engineering and Design, Israel, E-mail: iyad.suleiman@gmail.com*
- *Maha Arsla, Ph.D. is currently a lecturer in Sakhnin College, Israel, E-mail: maha.arslan@gmail.com*
- *Reda Alhajj, Full Professor in Computer Science in Calgary University, Canada, Email: rsalhajj@gmail.com*
- *Mick Ridley, Lecturer in Bradford University, UK, Email: M.J.Ridley@bradford.ac.uk.*

theory to transform practice to benefit learners. EDM has emerged as a research area in recent years for researchers all over the world from different and related research areas such as:

- Offline education try to transmit knowledge and skills based on face-to-face contact and also study psychologically on how humans learn. Psychometrics and statistical techniques have been applied to data like student behaviour/performance, curriculum, etc. that was gathered in classroom environments
- E-learning and Learning Management System (LMS). E-learning provides online instruction and LMS also provides communication, collaboration, administration and reporting tools [54]. Web Mining (WM) techniques have been applied to student data stored by these systems in log files and databases [70].
- Intelligent Tutoring (ITS) and Adaptive Educational Hypermedia System (AEHS) are an alternative to the just-put-it-on-the-web approach by trying to adapt teaching to the needs of each particular student [68]. Data Mining has been applied to data picked up by these systems, such as log files, user models, etc. [68].

The sources of information to be mined are heterogeneous. They include databases of the students' profile, log assessments of the user's interaction with the system, evaluation records, background knowledge, educational content, learning objects, student models, tutoring strategies, meta-data, federative teaching services, and many more repositories. Therefore, a sample of Educational Data Mining **(**EDM) applications is shown in this section according to the source of knowledge.

2.1 Student Modelling

Student models represent information about student's characteristics (e.g., student's knowledge, motivation, skills, personality, and learning preferences). An interesting EDM work oriented to student modelling is the comparison of student skill knowledge methods carried out by [4]. The study analyzes three methods for estimating students' current stage of skill mastery, such as: common conjunctive cognitive diagnosis model, sum-score method, and capability matrix.
Therefore, they try to estimate for a given topic the degree of skill achieved (e.g., complete, partial, none).

2.2 Tutoring

Tutoring corresponds to the traditional support that a human tutor offers to students to solve problems of a specific domain. This kind of functionality is fully implemented in intelligent tutoring systems (ITS). Regarding the application of DM in the tutoring field, the work achieved by [7] uses hints generated from historical data to develop logic proofs. Hints are outcome by a reinforcement learning technique based on Markov decision processes.
With reference to the framework stated by [32], it uses DM algorithms based on evolutionary computation to characterize dynamic learning processes and learning patterns for encouraging students' apprenticeship. The approach supports tutoring and collaboration functionalities to provide content that meet students' accessibility needs and preferences. The framework, also, pursues to match content to students' devices. These kinds of services are valuable for people with special abilities

2.3 Content

Content corresponds to the knowledge domain resources that are tailored to teach a lesson, record the students' behaviour, and evaluate students' apprenticeship. This resource is a kind of learning object that contains text, sound, image, video, virtual reality, animation, and many more multimedia options. An example of the DM application to content is given by [57]. They set a transfer model of the knowledge domain of related practice item-types using learning curves. The item-types mean a set of practice items that are alike. Such a model represents the pair wise knowledge component relationships between item-types in the domain.
Another DM contribution to the content line is the work fulfilled by [30]. They built a system to find, share and suggest the suitable modifications to improve the effectiveness of a course and its content. Their approach includes rule mining to discover valuable information through students' assessments like "if-then" recommendation rules. The system holds a collaborative recommender module to share and score the recommendation rules obtained by teachers and specialists in education with common profiles.

2.4 Assessment

The record of the user interaction with a Web-based Educational Systems (WBES) during each session is fulfilled by the assessment module. Based on the information stored, it is possible to supervise the behaviour, performance, outcomes, customs, preferences, and many more issues about: who is the student? And what has she/he been doing? As an instance of DM applications to assessment, there is a method for mining multiple-choice assessment data set by [46]. The method estimates similarity of the concepts given by multiple choice responses. As an outcome, a similarity matrix shows the distance between concepts in a lower-dimensional space. Such a view reveals the relative difficulty of concepts among the students. In addition, concepts are clustered, and unknown responses in the context of previously identified concepts are acknowledged. The method is used to answer questions related to the similarity of concepts and the difficulty of convincing students to modify an erroneous concept.
With the aim of focusing on the DM processes [58] stated a DM research line called "Process Mining". The line pursues the development of mining tools and techniques devoted to extract processes-related knowledge from event logs recorded by the system. One EDM application of process mining is devoted to analyze assessment data. The approach analyses assessments from recently organized online multiple-choice

tests. It, also, demonstrates the use of process discovery, conformance checking and performance analysis techniques.

2.5 Conclusions

As the Internet and World Wide Web are rapidly developing, the technologies that support the educational processes come to replace the traditional educational systems. More and more teachers provide their teaching material to their students through simple or more sophisticated electronic means and experts in various fields continually provide knowledge to the public, usually in the form of web pages.

According to [14], Adaptive and Intelligent Web-Based Educational Systems provide an alternative to the traditional 'just-put-it-on-the-Web' approach in the development of Web-based educational courseware. In their work Brusilovsky and Pyelo (2003) mention that Adaptive and Intelligent Web-Based Educational Systems attempt to be more adaptive by building a model of the goals, preferences and knowledge of each individual student, and by using this model throughout the interaction with the system in order to be more intelligent by incorporating and performing some activities which are traditionally executed by a human teacher – such as tutoring, assessing, or preparing corresponding content.

3 Poverty and Education

Through a combination of international development frameworks such as the
Millennium Development Goals (MDGs), the Education for All (EfA) goals and the World Fit for Children (WFfC) targets, countries are working towards a society in which all children will complete primary or basic education at a [37].

It is true that more children enter school, however, it is apparent that many of them are enrolling too late or too early, repeating grades, dropping out or failing to learn. It is gaining global support as a viable means to help young children reach their full developmental potential and engage in lifelong learning. School readiness is linked to improved academic outcomes in primary and secondary school and positive social and behavioural competencies in adulthood.

With respect to high school outcomes and academic achievement, the links to school readiness have also been established [72]. Data from several developing countries, including Brazil, Jamaica and the Philippines, indicate a strong association between early skills and later high school completion, controlling for a host of influencing factors such as family income and education [31].

According to a study by [8], "poor children who attended one year of preschool stayed in primary school 0.4 years longer than children who did not attend preschool. For each year of preschool, children have a 7-12 percent increase in potential lifetime income, with the larger increases gained by children from families whose parents had the least amount of schooling" [60]

The study by [79] from Latin American countries shows in Figure 1, that Cuba shows much better performance than other major Latin American countries. The Cuban results different from those of Chile, Brazil, Argentina, and Colombia because of the education system and the investment in mothers and children.

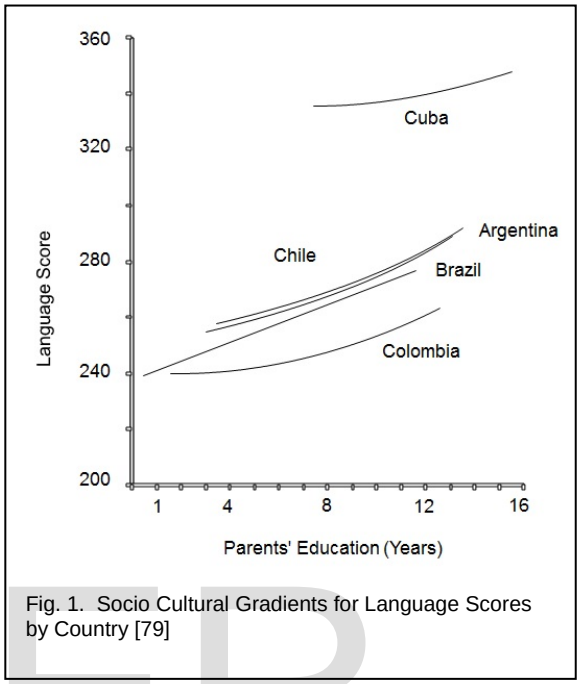

Fig. 1. Socio Cultural Gradients for Language Scores by Country [79]

4 Social Network Analysis

The Social Network Analysis deals with the analysis of the relationships that exist between entities in a social network. For instance, in a social network of people, the analysis can include who is friend with whom, who can influence which group of people, whom can have access to the information that goes through this network etc. Lately there has been a growing interest in this field, especially as to how it gets involved with knowledge discovery and data/web mining. For instance, analyzing the behaviour of users in online discussions or discover how users form communities and are affected by them are interesting works. Researchers analyze newsgroups by applying Social Network techniques and they interpret online communities by assigning roles to the members of the groups [28]. This is done by observing how people relate to each other in a graph-based model of post-reply relations. They notice that short discussion threads point out question-answer exchanges and longer threads indicate proper discussions. [36] analyze the Twitter's social network and the intentions of the associated users in order to understand the reason why people use such networks. They identify the communities that are formed, they categorize them into communities that create information, communities that receive information and communities that exist only because of friendship. They label the identified communities by the keywords that appear in the various posts.

In data analysis models which are used to predict future data trends are known as predictive analysis models. Classification or estimation algorithms are central in predictive analytics and are used in many areas of human endeavour, including (but not limited to) business and science. Examples of application areas from business include credit approval, medical diagnosis, performance prediction and selective marketing. Predictive models assess unlabeled samples to determine the value or value ranges of an attribute that a sample is likely to have [34]. With predictive analysis the validity of the classification result (and the true accuracy of the model) can be verified by waiting for the future event to happen. Though predictive accuracy is a critical aspect of models there are other facets that are equally important. We may require a model to show which of the predictor variables are most important in the dataset [75]. We may be interested in examining whether predictor variables interact or whether a simple model can result in good prediction. In the research, presented in this paper, we are interested in taking in account the structure of "social" relationships between the children in a predictive modelling dataset. In particular, we consider enriching the predictive modelling dataset with attributes that represent information about the structure of such relationships. Such attributes are based on concepts from social network analysis (SNA). In this paper we append attributes that correspond to some SNA centrality measures and then test the hypothesis that appending centrality measures improves the prediction accuracy. For the purpose of the paper, the dataset we use is a snapshot of a three year span, that, to some extent, encapsulates the temporal relationship of predictors to the target variable. Linoff (2004).

Social network analysis (SNA), which consists in generating patterns that allow, identifying the underlying interactions between users of different platforms, has been an area of high impact in the last years. The appearance of social networking services, such as Facebook or Twitter, has caused a renewed interest in this area, providing techniques for the development of market research using the activity of the users within those services.
However, SNA techniques do not just concentrate on social networks, but also focus on other fields, such as marketing (customer and supplier networks) or public safety, Krebs (2002). One of the fields in which they are also applied is education[64].

Thanks to SNA, it is possible to extract different parameters from the student activity in online courses, e.g., the students' level of cohesion, their degree of participation in forums, or the identification of the most influential ones. This kind of analyses might be helpful for teachers to understand their students' behaviour, and as a consequence, help them to get better results.

SNA is also useful for generating new data as attributes, which can be subsequently processed using data mining techniques to obtain student behaviour patterns. In the educational field, there is a well-defined area called educational data mining [68]. Building accurate performance and dropout predictors, which help teachers to prevent students from failing their subjects, is one of the main problems tackled in this area. For this purpose, classification techniques, by means of prediction models, are usually applied to uncover the students' behaviour, e.g., amount of time dedicated to accomplish certain tasks or activity in forums that results in a pass, a fail, or a dropout. For the issue of prediction, SNA provides a new useful framework that might improve the accuracy of those models.

In this paper, survey data was analyzed from the Life school for Creativity and Excellence and another 3 different schools for three consecutive academic years. In the data analyzed, SNA helps to uncover behaviour patterns and build models that predict the performance and dropouts of children accurately.

We propose a prediction model to evaluate the readiness of a child to start school based on the social factors mentioned above in addition to the computerized assessment results. In this work, data mining techniques were used, including clustering, classification, and social network analysis[9]. Due to the difference in school readiness assessment from one school to another, the classification model was built in a way that allows schools to modify the classifier to be used to add features that are used in the particular school.

5 Predicting School Readiness by Using Data Mining Techniques

5.1 EDM techniques

Recent years have shown a growing interest and concern in many countries about the problem of school failure and the determination of its main contributing factors. The great deal of research [3] has been done on identifying the factors that affect the low performance of students (school failure and dropout) at different educational levels (primary, secondary and higher) using the large amount of information that current computer can store in databases. All these data are a "gold mine" of valuable information about students. Identifying and finding useful information hidden in large databases is a difficult task [62]. A very promising solution to achieve this goal is the use of knowledge discovery in databases techniques or data mining in education, called educational data mining, EDM [69]. This new area of research focuses on the development of methods to better understand students and the settings in which they learn [68].
There are good examples of how to apply EDM techniques to create models that predict dropping out and student failure specifically, Kotsiantis, Patriarcheas, and Xenos (2010). These works have shown promising results with respect to those sociological, economic, or educational characteristics that may be more relevant in the prediction of low academic performance. It is also important to notice that most of the research on

the application of EDM to resolve the problems of student failure and drop-outs has been applied primarily to the specific case of higher education [39] and more specifically to online or distance education [45]. However, very little information about specific research on preschool, elementary and secondary education has been found, and what has been found uses only statistical methods, not DM techniques [5]. Although "Statistics and visualization" cannot formally be considered data mining, statistics can be often included as the starting point of any study [69]

There are several important differences and/or advantages between applying data mining and using statistical models [2]:

1. Data mining is a broad process that consists of several stages and includes many techniques, among them the statistics. This knowledge discovery process comprises the steps of pre-processing, the application of DM techniques and the evaluation and interpretation of the results.
2. Statistical techniques (data analysis) are often used as a quality criterion of the verisimilitude of the data given the model. DM uses a more direct approach, such as to use the percentage of well-classified data.
3. In statistics, the search is usually done by modelling based on a hill-climbing algorithm in combination with a verisimilitude ratio test-based hypothesis. DM is often used as a meta-heuristics search.
4. DM is aimed at working with very large amounts of data (millions and billions). The statistics alone do not usually work well in large databases with high dimensionality.

This paper proposes to predict child readiness at pre-school in elementary education by using DM. In fact, the goal is to detect the factors that most influence child readiness in preschool by using association rules mining, clustering and classification techniques. Also different techniques of DM have been used because the problem is complex, i.e., the data is characterized by high dimensionality (there are many factors that can influence) and it is often highly unbalanced (the majority of children pass and too few fail). The final objective is to detect as early as possible the children who show these factors in order to provide some type of assistance for trying to avoid and/or reduce school failure.

5.2 Method

The method proposed in this paper for predicting the school readiness of children belongs to the process of Knowledge Discovery and Data Mining (see Fig. 2). The main stages of the method are:

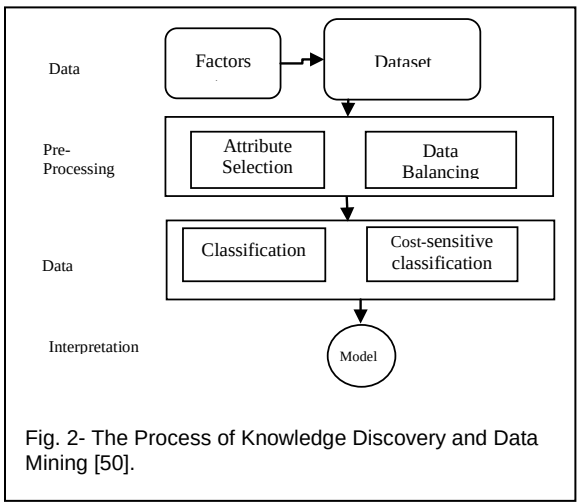

Fig. 2- The Process of Knowledge Discovery and Data Mining [50].

1. *Data pre-processing* is an important step in the data mining process. Data-gathering methods are often loosely controlled, resulting in out-of-range values (e.g., Income: −100), impossible data combinations (e.g., Sex: Male, Pregnant: Yes), missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis [61].
2. *Data mining*. At this stage, DM algorithms are applied to predict child readiness like a frequent pattern mining, clustering or classification problem. To do this task, it is proposed to use:
3. *Frequent pattern mining algorithm*, e.g., Apriori, was applied to find groups of students sharing same characteristics. This is achieved by considering students as items and characteristics of students as transactions. Then frequent sets of students are found by analyzing their common characteristics. Every frequent set of students with cardinality larger than one reveals some interesting information about the students inside the set. The support of the set reflects the strength of the relationship between the students in the set by considering their characteristics.
4. *Clustering of students using hierarchical clustering or k-means,* k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster, this will allow us to investigate each group of students forming one cluster and their distribution within the cluster. Students closer to the centre of the cluster are more interesting and solid entities inside the cluster than those closer to the boundary of the cluster. The study also investigtes how the outcome from the frequent pattern mining process matches the outcome from the clustering process. It is anticipated that students who end up in the same cluster are mostly together in the same frequent set of students.
5. *Classification algorithms based on splitting the data into training and test sets*. The training data will be used for building the classifier model and the test set will be used to evaluate the model. This method has two basic drawbacks:

(1) In problems where we have a sparse dataset we may not be able to afford the "luxury" of setting aside a portion of the dataset for testing (2) Since it is a single train-and-test experiment, the estimate of error rate will be misleading if we happen to get an "unfortunate" split. The limita-

tions of this method can be overcome with a family of resampling methods at the expense of more computations, like: Cross Validation, and Bootstrap. 10-fold cross validation is applied where the data is split into 10 disjoint subsets. Nine subsets form the training set: used to train the classifier, and the 10-th subset is used as the test set: used to estimate the error rate of the trained classifier.

The outcome from the frequent pattern mining and clustering will provide excellent input for constructing the social network of the students. This is essential because students who end up in the same frequent set or in the same cluster are likely to be similar and hence linked together. The weight of the link is determined based on the collective support of the sets in which the two students exist together combined with the value obtained from the distance separating the two students from each other and from the centroid of their cluster.

6. *Interpretation*. At this stage, the obtained models are analyzed to detect child readiness. To achieve this, the factors that appear and how they are related are considered and interpreted. Students in the same frequent set or cluster are expected to show the same trend and level of readiness. The degree of confidence in this result is determined by the support of the set of students produced by the frequent pattern mining process or based on the distance of the two students from the centroid of their cluster. The classifier model will support this result by either producing the same class for both students or not. However, in case the classifier does not produce same class for both students then the interpretation will be based on the frequent set and cluster analysis to understand why the two students could not end in the same class. In other words, the support of the dataset and the distance within the cluster will lead to good interpretation of how far away the student will be classified, i.e., are they very close to being in the same class or not.

The next step is a description of a case of study with real data from Arab children in order to show the utility of the proposed method.

5.3 Data Gathering

School failure of students is also known as the "one thousand factors problem" [50], due to the large amount of risk factors or characteristics of the students that can influence school failure, such as demographics, cultural, social, family, or educational background, socioeconomic status, psychological profile, and study progress.

In this paper, information has been used from pre-school children enrolled in *Life school for Creativity and Excellence* and three other schools for three consecutive academic years, Sep 2008- June 2013. The information used was only about pre-school children, where most children are between the ages of 5 and 6, as this is the year for moving from pre-school to 1st grade. All the information used in this study has been gathered from three different sources during the aforementioned period:

1. A general survey was designed and administered to all children in the middle of the year. Its purpose was to obtain personal and family information to identify some important factors that could affect school performance.
2. From a specific survey (Teacher questionnaire) which is completed when the children register for admission to kindergarten and pre-school in the school and also the results of the assessment conducted by the Kindergarten/Preschool teacher in the beginning of second semester (Feb-Mar).
3. The Teacher provides the scores/evaluations obtained by children in all subjects of the pre-school in the end of the academic year.

In Table 1, all the used variables in this study are shown grouped by data source.

5.4 Data Pre-Processing

Before applying DM algorithms it is necessary to carry out some pre-processing tasks such as cleaning, integration, discretization and variable transformation, Márquez-Vera (2013). It must be pointed out that a very important task in this paper was data pre-processing, due to the quality and reliability of available information, which directly affects the results obtained. In fact, some specific pre-processing tasks were applied to prepare all the previously described data so that the classification task could be carried out correctly. Firstly, all available data were integrated into a single dataset. During this process those children without 100% complete information were eliminated.

TABLE 1
VARIABLES USED AND INFORMATION SOURCES

Source	Variable
General survey	Classroom/group, number of friends, parental encouragement for study, religion, the type of personality, having a physical disability, suffering a critical illness, family income level, mother' s level of education, father' s level of education, number of brothers/sisters, position as the oldest/middle/youngest child, [Social factors]: *number of Peers in Class (Good, Average, Poor), number of Peers in neighbourhood (Good, Average, Poor)*, living in a large city, number of years living in the city, transport method used to go school, distance to the school, level of attendance during classes, interest in the subjects, level of difficulty of the subjects, level of motivation, quality of school infrastructure, level of teacher' s concern for the welfare of each student.
Specific survey	Academic year, Age, sex, previous school, type of school, mother's occupation,

(Teacher questionnaire)	father's occupation, number of family members, limitations for doing exercises, frequency of exercises, time spent doing exercises, scores obtained in Count Balloons, Count Balloon Strings, Identify the Number, Amount, Digit Matching, More or Less, Addition & Subtraction, Choose the Form, Magic Circle, Incomplete Shadow, Triangles, Analogy, Remember the Location, Sequence of Events, Identifying Faces, Hand Movements, Picture Selection, Picture Recognition, Series of Pictures, Series of Numbers, Ba[illegible]l Digital Series, Sound Units, Identi[illegible], Match Rhyming Words, Opening Sound, Closing Sound
Department of school services (Evaluation)	Score in Arithmetic Readiness, score in Cognitive Development, score in Language Development, score in Phonological Awareness, score in Chess, score in Arts, and score in Computer skills.

Fig 1

All children who did not answer one of the specific surveys were excluded. Some modifications were also made to the values of some attributes.

A new attribute of the age of each student in years was created using the day, month, and year of birth of each student. Furthermore, the continuous variables were transformed into discrete variables, which provide a much more comprehensible view of the data. For example, the numerical values of the scores obtained by children in each subject were changed to categorical values in the following way:

Excellent: score between 95 and 100; Very good: score between 85 and 94; Good: score between 75 and 84; Regular: score between 65 and 74; Sufficient: score between 60 and 64; Poor: between 40 and 59; Very poor: less than 40 and Not presented.

Then, all the information was integrated in a single dataset and it was saved in the .ARFF format

6 Data Formatting

Data mining is an integral part of Knowledge Discovery in Databases (KDD), which is the overall process of converting a series of transformation steps, from data pre-processing to post-processing of data mining results. The data pre-processing has to do with gathering or collection of data, and data cleaning through data transformation. During data selection, the relevant data is gathered. Once the data has been assembled, its quality must be verified.

Incomplete (lacking certain attributes of interest, or containing only aggregate data), noisy (containing errors, or outlier values that deviate from expected), and inconsistent (for example, discrepancies in the codes used to categorize items) data are common. Data cleaning routines attempt to clean the data by filling in missing values; smoothing noisy data, identifying or removing outliers, and resolving inconsistencies. Finally, the cleaned data are transformed into a format suitable for data mining.

The data gathering process for this study involves the collection of the raw data about the children from Table 2.

According to the training data set, seven distinguishing features associated with each child (row): father's education, father's job, mother's education, mother's job, family size, child position as well as siblings and friends which are combined into one feature. These features represent the relationship between Socio-Economic Satus (SES) and school readiness as demonstrated in the literature [72]. These features represent four different aspects: parent's education, parent's job, family composition and peers including siblings and friends.

By analyzing the data it can give an idea on how each aspect, let's say parent's education may affect child's readiness regardless of other feature values and so on. For the parent's education, all children were classified in a training set based on their parent's education levels. Finding that parent's education provided in the data has six possible values, Primary, Secondary, 1st_degree, 2nd_degree, MD, PhD

Based on parent's jobs, jobs were classified into three classes: "UnEmployed", "Private" and "Government. The third aspect involves considering the family, including family size and child position within the family.

The fourth aspect considers the peers factor, including siblings and friends where friends cover both friends at school and in the local community at home. For every child, it was decided to explore his/her peers and check their achievement level by dividing them into four groups: good class peers, weak class peers, good neighbourhood peers and weak neighbourhood peers.

In accordance with the attribute's main pedagogical impact from the expert's points of view, respective classification attributes were defined as follow:

- Mother educational qualification whose labels are :{Primary, Secondary, 1st_degree, 2nd_degree, MD, PhD}
- Father educational qualification whose labels are: {Primary, Secondary, 1st_degree, 2nd_degree, MD, PhD}
- Mother occupation whose labels are: {UnEmployed, Private, Government}
- Father occupation whose labels are: {UnEmployed, Private, Government}
- Family size whose labels are: {='Big' if >4 members, ='Small' if <=4 members}
- Child position whose labels are: {='Late' if after 2nd child, ='Top' if before 2nd child}
- Good Class Peers whose labels are: {='Good' if<=2 peers, ='Weak' if>=2 peers}
- Weak Class Peers whose labels are: {='Good' if<=2 peers, ='Weak' if>=2 peers }
- Good Neighbourhood Peers whose labels are: {='Good' if<=2 peers, ='Weak' if>=2 peers}
- Weak Neighbourhood Peers whose labels are: {='Good' if<=2 peers, ='Weak' if>=2 peers}

- Ready4School whose labels are: {='NotReady' if not ready, ='Ready'=if ready}

7 Predictors of School Readiness and Social-Emotional Competence

Most research on school readiness has focused on family risk factors, and the ways that multiple risk factors in families negatively affect school readiness in children [23]. Families that experience economic, social, and/or psychological hardship, and have few resources to cope with these tend to experience higher rates of school "un-readiness" than do more advantaged families [23].

There are some researchers who argue that the children's home environments do not provide the best support for the early development of their school readiness skills, especially in families who are low-income and come from culturally diverse backgrounds [26]. [48] used an integrative theoretical model of child development formulated specifically for understanding development among children of colour.

Presently, researchers are expanding how to understand the ecological influences on the development of academic readiness skills, including both family and school-related factors [19]. Unfortunately, researchers still cannot determine which aspects of socioeconomic conditions (e.g., income, parental occupation) contribute to the improvement of a child's readiness for school [72]. In addition, the reader must be cautious of other researches who provide estimates of how much different factors contribute to the overall readiness gap. Given that these factors are highly correlated with one another, any one factor can pick up the effects of others, therefore making it extremely difficult to look at one factor individually.

The next section describes the factors that were included in this paper as predictors of school readiness and social-emotional competence.

TABLE 2
CHILD DATA FORMAT

#	Variable Name	Variable description/format	Variable Type
1	Age on entry	Students age on admission Continuous	Continuous
2	Gender	Male or Female Categorical	Categorical
3	Social class	Upper, Middle, Lower	Categorical
4	Mother's educational qualification	Primary, SSCE, 1st degree, 2nd degree, PhD	Categorical
5	Father's educational qualification	Primary, SSCE, 1st degree, 2nd degree, PhD	Categorical
6	Marital status of parents	Married, Divorced, Separated, Widowed	Categorical
7	Parent's relationship	Healthy, Problematic	Categorical
8	Mother's occupation	Government worker, Private, Self employed	Categorical
9	Father's occupation	Government worker, Private, Self employed	Categorical
10	Family size	Total number of children in family and parents	Continuous
11	Child's position in the family	1st born, last born, only child, others	Categorical
12	Type of kindergarten attended	Private, Missionary school, Public	Categorical
13	Location of kindergarten	Rural, Semi-Urban, Urban	Categorical
14	Residence location	Rural, Semi-Urban, Urban	Categorical
15	Class Peers with level Good	Number of Peers in Class with grade level (70-100)	Continuous
16	Class Peers with level Weak	Number of Peers in Class with grade level (1-69)	Continuous
17	Neighbourhood Peers with level Good	Number of Peers in neighbourhood with grade level (70-100)	Continuous
18	Neighbourhood Peers with level Weak	Number of Peers in neighbourhood with grade level (1-69)	Continuous
19	Arithmetic Readiness score	Total Arithmetic Readiness result score (0-100)	Continuous
20	Cognitive Development score	Total Cognitive Development result score (0-100)	Continuous
21	Language Development score	Total Language Development result score (0-100)	Continuous
22	Phonological Awareness score	Total Phonological Awareness result score (0-100)	Continuous
23	Chess evaluation	Very Good, Good, Satisfying, Weak	Categorical
24	Arts evaluation	Very Good, Good, Satisfying, Weak	Categorical
25	Music evaluation	Very Good, Good, Satisfying, Weak	Categorical
26	Computer skills evaluation	Very Good, Good, Satisfying, Weak	Categorical
27	Science evaluation	Very Good, Good, Satisfying, Weak	Categorical
28	Ready4School	Ready, Not Ready	Categorical

8 Socio-demographic Variables

8.1 Socioeconomic Status or Income

The literature suggests that income matters more for preschoolers than for older children and much more for poor children than for children from more economically advantaged situations [22]. Accounting studies find that differences in SES explain about half a standard deviation of the initial achievement gaps [65].

Family SES appears to explain a great amount of variance of racial and ethnic gaps in school readiness [72]. Family SES is important for school readiness because it underlies many of the factors that affect school readiness [72]. Life for a family in a low socioeconomic household is very different than for a family living in a more advantageous situation [22]. The first family may provide a lower quality home environment for a child and provide fewer learning opportunities in the home or in an outside lower-quality child care [22]. The second family, however, may be the total opposite, where parents read to their children, visit museums, and engage in conversations.

In families with a low SES, parents are less likely to read or talk to their children than are parents in a more economically advantaged situation. The results of these behaviours are associated with school readiness given the relationship between school readiness and socioeconomic conditions and parenting

behaviours [72]. Differences such as these suggest that SES plays a significant role in school readiness and why it is necessary to take it into account in studies of children's school readiness.

Studies have found a relationship between SES and school readiness. In an analysis of the data of the 1998 Early Childhood Longitudinal Study, ECLS-K [55], [18] found that SES was related to proficiency across all reading tasks, where children in higher SES groups were more likely to be proficient than children in lower SES groups. SES was related to proficiency in all mathematics tasks, where children in higher SES groups were more likely to be proficient than were children in lower SES groups.

A relationship between SES and social-emotional competence has been demonstrated in the literature. Low-income children are at the highest risk of developing emotional and behavioural difficulties [11], the poverty status and SES are significant predictors of children's early language skills and academic achievement, and social competence [51].

8.2 Family Size

Head Start children tend to have mothers who come from large families and households that are less likely to have had either an adult male or an adult female working when the mother was 14 [20].

Crowded home environments have been associated with disparities in children's social functioning, vocabulary growth rates, and cognitive abilities [35]. Parents have also been rated as being less responsive to their children when compared to those who were living in less crowded homes [77]. The degree of stress associated with high density home environments has been shown to be negatively correlated with the frequency of parent to child speech [77], also the family size was negatively associated with children's literacy interest [26], such that children who engaged in literacy-related behaviours had smaller families. It was found that children from small families (one sibling or less) had higher scores on expressive language skills than children from large families (three siblings or more) [74]. In addition, also it was found that family size of four or more children was a risk factor in poor cognitive and social emotional development in preschool children [73].

It looked as if the number of adults and children living in the household is a predictor of school readiness and social-emotional competence. It was hypothesized that children from larger families would have lower school readiness and social-emotional competence.

8.3 Education of the Caregiver

The most studied form of human capital is formal schooling [22]. Research has shown that parental education plays a role in determining a child's educational experience [59]. In addition, children who have highly educated parents typically obtain higher scores on cognitive and academic achievement tests than do children of parents who have less education [22]. Other researchers have stated that children from low education parents tend to perform less adequately in cognitive skills than children from better educated parents [67]. In an analysis of the data of the 1998 Early Childhood Longitudinal Study, ECLS-K: [55], It was found that having parents with less education put a student at-risk for school failure [18]. It was found that maternal education was associated with academic achievement and successful grade completion [29].

In addition to these studies, other researchers have supported parental education's role in school readiness [82], also the level of maternal education was strongly related to each of the literacy-numeracy accomplishments.

8.4 Working Caregiver

The research on having a caregiver that works as a predictor of school readiness and social-emotional competence has been little studied and mixed.

Head Start children have been found to be less likely to have mothers that work [20]. Research indicated that the research on the effects of occupation on young children is sparse [22], [66].

Found that maternal employment increased the likelihood that children would experience "high stable" environments. Children in "high stable" environments had higher scores in school readiness than children in "low rise" environments. [60] indicates that given the financial benefits of working, mothers who are employed might be better able to invest in stimulating learning materials and engage in educational activities (e.g., visiting a museum) that may in turn promote learning in their children. Contrary to [66] findings, it was found that maternal employment by the ninth month was found to be linked to lower school readiness scores at 36 months. The effects were stronger when mothers were working 30 hours or more a week [12].

8.5 Peer Interactions

Peer interactions are viewed as a developmental context for learning. Through interactions with their peers, young children practice the important skills necessary for competent social and academic adjustment to school[52]. In the preschool classroom children use their peer play interactions to work through more complicated academic material presented during instructional periods. Also, peer play in preschool is one context where children learn and practice the new demands and expectations of the school [24], [25]. Thus peer interactions can be a positive force in a child's life that help them develop the necessary skills to adapt to more advanced social and academic challenges in preschool classrooms.

It is also known that peer interactions are related to children's adjustment to school [42]. Children view friendships as a major concern when transitioning into new schools [43]. Peer interactions in elementary school have far-reaching effects, Peer Interactions and School Readiness Peer interactions are viewed as a developmental context for learning. Through interactions with their peers, young children practice the important skills necessary for competent social and academic adjustment to school [52]. In the preschool classroom children use their peer play interactions to work through more complicated academic material presented during instructional peri-

ods. Also, peer play in preschool is one context where children learn and practice the new demands and expectations of the school [24], [25]. Thus peer interactions can be a positive force in a child's life that help them develop the necessary skills to adapt to more advanced social and academic challenges in preschool classrooms.

It is also known that peer interactions are related to children's adjustment to school [42]. Children view friendships as a major concern when transitioning into new schools [43]. Peer interactions in elementary school have far-reaching effects, aggression and victimization, relational aggression and victimization, displayed and received pro-social behaviours, and school readiness will shed new light on the links between social-emotional development and children's early school success.

9 Apply Data Mining and Interpret Results

For this stage, WEKA was used (Waikato Environment for Knowledge Analysis) [80]; it is an open source package which provides data mining algorithms for clustering, classification, and association. In this section, for each algorithm used in the study, the test characteristic and results obtained are shown (see appendix 9). These results can be presented in the form of tables or graphs.

9.1 Association Algorithms

For the association rules generation, Apriori algorithm was executed [1]. For this algorithm, A generation of 100 rules were determined, based on the following parameters: a minimum support of 0.3 and minimum confidence of 0.9 as parameters, which have been set arbitrarily.

A set of IF-THEN rules were obtained from the algorithms. After an analysis, rules that were base on irrelevant information were eliminated.

TABLE 3
SOME OF THE BEST RULES OBTAINED WITH THE APRIORI ALGORITHM

Reliability	Rules—Generated	Rules—Interpretation
1.00	Family_size=0; Father_occupation=1 ==> Ready4School=1	Small family, father has a private job.
1.00	Mother_occupation=1; Family_size=0; Father_occupation=1 ==> Ready4School=1	Small family, father and mother have a private job.
1.00	Father_educational_qualification=0 ; Child_position=1; Father_occupation=1 ==> Ready4School=1	The father has a private job with primary education and a middle child in the family.
1.00	Mother_occupation=1; Good_Class_Peers=0; Father_occupation=1 ==> Ready4School=1	The father and mother have a private job and at the most 2 good class peers.
1.00	Mother_occupation=1 ; Good_Neighbourhood_Peers=0; Father_occupation=1 ==> Ready4School=1	The father and mother have a private job and at the most 2 good neighbourhood peers.
1.00	Child_position=1; Good_Class_Peers=0; Father_occupation=1 ==> Ready4School=1	The father has a private job and at the most 2 good class peers and a middle child in the family.
1.00	Father_educational_qualification=0; Mother_occupation=1; Child_position=1; Father_occupation=1 ==> Ready4School=1	The father and mother have a private job and the father with primary education and a middle child in the family.
1.00	Child_position=1 ; Good_Neighbourhood_Peers=0; Father_occupation=1 ==> Ready4School=1	The father has a private job with a middle child and at the most 2 good neighbourhood peers.
1.00	Mother_occupation=1; Weak_Neighbourhood_Peers=0; Father_occupation=1 ==> Ready4School=1	The father and mother have a private job with at the most 2 weak neighbourhood peers.
1.00	Father_educational_qualification=0 ; Good_Class_Peers=0; Father_occupation=1 ==> Ready4School=1	The father has a private job with primary education with at the most 2 good class peers.
1.00	Mother_educational_qualification=0; Mother_occupation=1; Child_position=1; Father_occupation=1 ==> Ready4School=1	The father and mother have a private job and a mother's secondary education with a middle child in the family.

9.2 The APriori algorithm

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database.

9.3 Clustering Algorithms

For clustering testing, the following algorithms were used: SimpleKmeans [83] and EM (Expectation Maximization), [84]. In each algorithm, the number of clusters was calibrated to generate the greater amount of clusters having mutually exclusive attributes.

9.4 The k-means algorithm

The *k*-means algorithm is a simple, straightforward algorithm to assign instances to clusters. Each cluster is defined by a cluster centroid, and instances belong to the cluster for which their Euclidian distance to the centroid is the smallest. For each cluster a new centroid is found by taking the average over the cluster instances, which may lead to shifts of instances between clusters. This iterative process ends when the centroids stop changing. In Table 35 and 36, some clusters obtained are presented.

TABLE 4
CLUSTERING RESULTS - SIMPLEKMEANS (FULL TRAINING DATA)

		Cluster#				
Attribute	Full Data	0	1	**2**	3	4
	(306)	(84)	(73)	**(87)**	(31)	(31)
Mother_educational_qualification	Secondary	Secondary	1st_degree	**Secondary**	1st_degree	Secondary
Father_educational_qualification	Secondary	Secondary	1st_degree	**Secondary**	Secondary	Secondary
Mother_occupation	Private	Private	Private	**Private**	Private	Government
Father_occupation	Private	Private	Private	**Private**	Private	Private
Family_size	4.7026	4.9643	4.6849	**4.5862**	4.3548	4.7097
Child_position	1.9346	2.2738	1.7671	**1.8621**	1.5806	1.9677
Good_Class_Peers	2.9641	4.0476	2.9726	**1.9195**	3.2903	2.6129
Weak_Class_Peers	2.8203	3.0357	2.9041	**2.5517**	3.1290	2.4839
Good_Neighbourhood_Peers	2.8725	2.7976	2.8630	**3.0920**	2.7097	2.6452
Weak_Neighbourhood_Peers	2.8268	3.2857	2.8082	**2.1954**	2.8065	3.4194
Ready4School	Yes	Yes	Yes	**Yes**	Yes	Yes

TABLE 5
CLUSTERING RESULTS - SIMPLEKMEANS (PERCENTAGE SPLIT)

		Cluster#				
Attribute	Full Data	0	1	2	3	**4**
	(201)	(39)	(35)	(38)	(37)	**(52)**
Mother_educational_qualification	S.School	S.School	S.School	S.School	S.School	1st_degree
Father_educational_qualification	S.School	S.School	S.School	S.School	S.School	1st_degree
Mother_occupation	Private	Private	Private	Private	Private	Private
Father_occupation	Private	Private	Private	Private	Private	Private
Family_size	4.7164	4.6667	4.5429	4.8158	4.7297	4.7885
Child_position	1.9353	1.7949	1.9429	2.1316	2	1.8462
Good_Class_Peers	3.0149	2.7692	3.1143	2.6053	3.6757	2.9615
Weak_Class_Peers	2.8358	2.9487	1.9143	1.9474	4.1892	3.0577
Good_Neighbourhood_Peers	2.9701	4.2821	3.9429	1.7895	1.7568	3.0577
Weak_Neighbourhood_Peers	2.7015	1.9231	4.2286	1.5789	3.2973	**2.6538**
Ready4School	Yes	Yes	Yes	Yes	Yes	**Yes**

In table 38 and table 39, the clusters 2 and 4 are frequent, then all of their subsets must also be frequent, the other item sets (clusters) are infrequent then all their supersets must also be infrequent [47].

TABLE 6
CLUSTERING RESULTS - EM (FULL TRAINING DATA)

	Cluster			
	0	1	2	3
	(0.28)	**(0.55)**	(0.03)	(0.14)
Mother_educational_qualification				
Primary	1.0814	1.0554	2.9917	4.8716
Secondary	3.2824	**153.2146**	2.4290	38.0740
1st_degree	79.6616	15.3416	6.9338	3.0630
2nd_degree	3.4803	1.3248	1.0118	1.1831
MD	2.4575	1.0126	1.5276	1.0023
PhD	1.0180	1.0016	1.0003	1.9801
[total]	90.9811	172.9506	15.8943	50.1740
Father_educational_qualification				
Primary	1.0821	1.0583	2.9918	5.8678
Secondary	16.9678	**154.3899**	6.3541	37.2882
1st_degree	65.7893	14.1211	3.2292	2.8604
2nd_degree	1.5708	1.2655	1.0100	2.1537
MD	4.5712	1.1158	1.3092	1.0038
PhD	1.0000	1.0000	1.0000	1.0000
[total]	90.9811	172.9506	15.8943	50.1740
Mother_occupation				
UnEmployed	6.8854	20.8486	2.7372	6.5288
Private	51.9359	**136.8048**	1.5246	34.7347
Government	29.1598	12.2972	8.6324	5.9106
[total]	87.9811	169.9506	12.8943	47.174
Father_occupation				
UnEmployed	1.0022	1.9410	1.0030	1.0538
Private	75.8401	**153.3461**	9.9799	40.8338
Government	11.1388	14.6634	1.9113	5.2864
[total]	87.9811	169.9506	12.8943	47.1740
Family_size				
Big	60.3667	**94.1779**	1.3863	45.0691
Small	26.6144	74.7727	10.5080	1.1049
[total]	86.9811	168.9506	11.8943	46.174
Child_position				
Late	11.5188	4.1964	1.0385	40.2463
Top	75.4623	**164.7542**	10.8557	5.9277
[total]	86.9811	168.9506	11.8943	46.1740
Good_Class_Peers				
Good	34.1558	67.3908	5.5991	13.8542
Weak	52.8254	**101.5597**	6.2951	32.3198
[total]	86.9811	168.9506	11.8943	46.174
Weak_Class_Peers				
Good	36.6143	**85.5786**	1.2019	22.6051
Weak	50.3668	83.3719	10.6924	23.5689
[total]	86.9811	168.9506	11.8943	46.174
Good_ Neighbourhood _Peers				
Good	38.6347	67.2160	8.3640	21.7853
Weak	48.3464	**101.7345**	3.5303	24.3887
[total]	86.9811	168.9506	11.8943	46.1740
Weak_ Neighbourhood _Peers				
Good	38.9568	78.6484	4.6824	22.7124
Weak	48.0243	**90.3021**	7.2119	23.4616
[total]	86.9811	168.9506	11.8943	46.174
Ready4School				
NotReady	6.2312	19.2632	5.2429	5.2627
Ready	80.7499	**149.6873**	6.6514	40.9113
[total]	86.9811	168.9506	11.8943	46.1740

TABLE 7
CLUSTERING RESULTS - EM (PERCENTAGE SPLIT)

	Cluster=0	1	2	3
	(0.67)	(0.18)	(0.14)	(0.01)
Mother_educational_qualification				
Primary	1.0014	1.0048	1.0009	1.9929
Secondary	**118.7411**	2.5448	3.6229	1.0912
1st_degree	17.5305	32.1962	26.2287	1.0447
2nd_degree	1.0682	2.8270	1.1051	1.9997
MD	1.0190	2.1451	1.8354	1.0005
PhD	1.0000	1.0000	1.0000	1.0000
[total]	140.3602	41.7178	34.7929	8.1291
Father_educational_qualification				
Primary	1.0014	1.0048	1.0009	1.9929
Secondary	**125.1342**	4.8114	11.9522	1.1021
1st_degree	11.1159	32.1136	16.6646	1.1059
2nd_degree	1.0169	1.0269	1.0293	1.9269
MD	1.0918	1.7610	3.1459	1.0012
PhD	1.0000	1.0000	1.0000	1.0000
[total]	140.3602	41.7178	34.7929	8.1291
Mother_occupation				
UnEmployed	11.0942	3.4893	4.3554	2.0611
Private	**113.1045**	23.4636	15.4609	1.9710
Government	13.1615	11.7650	11.9766	1.0969
[total]	137.3602	38.7178	31.7929	5.1291
Father_occupation				
UnEmployed	1.9509	1.0013	1.0385	1.0094
Private	**122.6440**	30.0981	28.2693	2.9886
Government	12.7653	7.6184	2.4851	1.1311
[total]	137.3602	38.7178	31.7929	5.1291
Family_size				
Big	**90.4762**	33.3225	9.0878	3.1135
Small	45.884	4.3953	21.7051	1.0156
[total]	136.3602	37.7178	30.7929	4.1291
Child_position				
Late	22.4184	9.2881	1.2288	2.0647
Top	**113.9418**	28.4297	29.5641	2.0643
[total]	136.3602	37.7178	30.7929	4.1291
Good_Class_Peers				
Good	47.9633	9.1317	18.8523	3.0528
Weak	**88.3969**	28.5862	11.9406	1.0762
[total]	136.3602	37.7178	30.7929	4.1291
Weak_Class_Peers				
Good	**70.3994**	16.4449	7.0645	3.0912
Weak	65.9608	21.273	23.7284	1.0378
[total]	136.3602	37.7178	30.7929	4.1291
Good_ Neighbourhood _Peers				
Good	53.0704	15.6327	14.2677	2.0292
Weak	**83.2898**	22.0851	16.5252	2.0999
[total]	136.3602	37.7178	30.7929	4.1291
Weak_ Neighbourhood _Peers				
Good	67.2351	16.4299	19.2255	3.1095
Weak	**69.1251**	21.288	11.5674	1.0196
[total]	136.3602	37.7178	30.7929	4.1291
Ready4School				
NotReady	17.2763	2.5023	7.2169	2.0045
Ready	**119.0839**	35.2156	23.5760	2.1245
[total]	136.3602	37.7178	30.7929	4.1291

9.5 The Expectation-Maximisation (EM) algorithm

The EM algorithm is a probabilistic clustering algorithm. Each cluster is defined by probabilities for instances to have certain values for their attributes, and a probability for instances to reside in the cluster. For numerical values it consists of a mean value and a standard deviation for each attribute value, for discrete values it consists of a probability for each attribute value.

The EM clustering scheme generates probabilistic descriptions of the clusters in terms of mean and standard deviation for the numeric attributes and value counts (incremented by 1 and modified with a small value to avoid zero probabilities) - for

the nominal ones. That shows the given instance belongs to each cluster with some probability. The overall likelihood is a measure of the "goodness" of the clustering and increases at each iteration of the EM algorithm. The larger this quantity, the better the model fits the data. Increasing the number of clusters normally increases the likelihood, but may lead to overfitting.

In the full training data mode the rule generated is the following:
Mother_educational_qualification= Secondary AND Father_educational_qualification= Secondary AND Mother_occupation= Private AND Father_occupation= Private AND Family_size= Big AND Child_position= Top AND Good_Class_Peers= Weak AND Weak_Class_Peers= Good AND Good_ Neighbourhood _Peers= Weak AND
Weak_ Neighbourhood _Peers= Weak THEN Ready

This above rule says the ready for school is affected by mother and father qualification and occupation and also the child position in the family, the rest of the parameters do not a have a strong relation to the readiness for school.
In the percentage split with 66% training mode the rule generated is the following:
Mother_educational_qualification= Secondary AND Father_educational_qualification= Secondary AND
Mother_occupation= Private AND Father_occupation= Private AND
Family_size= Big AND Child_position= Top AND Good_Class_Peers= Weak AND Weak_Class_Peers= Good AND Good_ Neighbourhood _Peers= Weak AND Weak_ Neighbourhood _Peers= Weak THEN Ready

The above rule is identical to the previous rule and same conditions will yield to readiness for school.

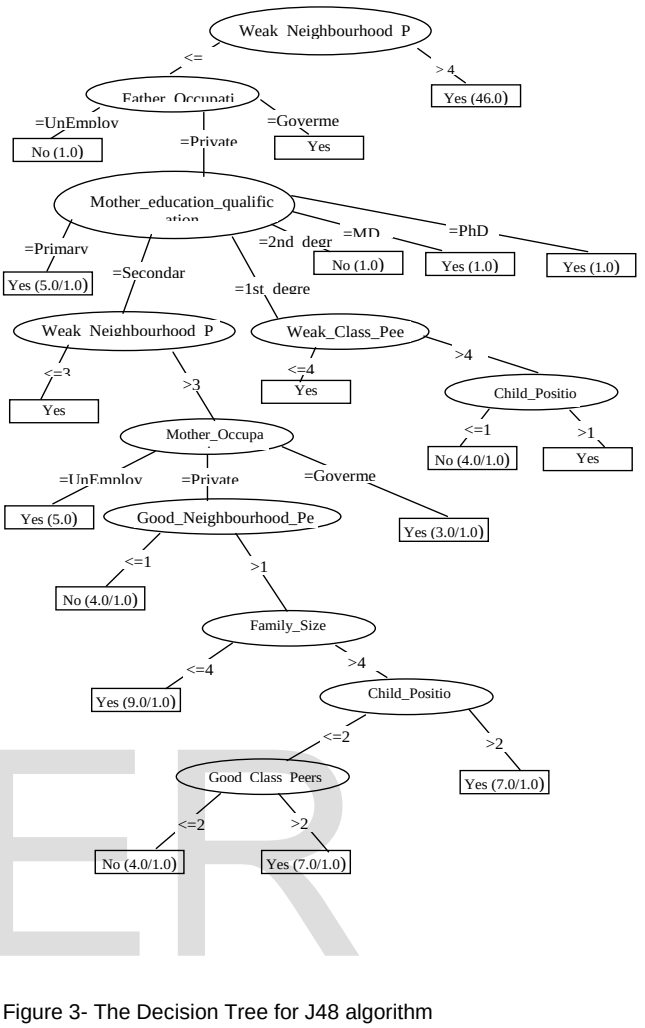

Figure 3- The Decision Tree for J48 algorithm

8.6 CLASSIFICATION ALGORITHMS

Some of the attributes that define the clusters were considered as a class. This is achieved using ID3 (induction decision trees) [63] and J48 algorithm [81]. These tests are intended to verify the effectiveness in the classification rules generation from both systems and thus provide corroboration if rules are similar. Various tests were verified with ID3 and J48 algorithms with the already mentioned dataset.
A set of IF-THEN-ELSE rules were obtained from the algorithms. After an analysis, rules with irrelevant information were eliminated. Tables 42 and 43 show some of the best rules obtained.

9.6 The J48 algorithm

A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision.
Decision tree are commonly used for gaining information for the purpose of decision-making. Decision tree starts with a

root node on which it is for users to take actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome.

TABLE 8
SOME OF THE BEST RULES OBTAINED WITH THE ID3 ALGORITHM

	Rules—Generated	Rules—Interpretation
1	Father_occupation = Private AND Father_educational_qualification = Secondary AND Mother_occupation = Private AND Mother_educational_qualification = Secondary AND Weak_Neighbourhood_Peers = Good AND Good_Neighbourhood_Peers = Good Family_size = Big THEN Ready	The readiness of the child is based here on parental secondary education, parental private job level, and good neighbourhood peers and big family size, respectively.
2	Father_occupation = Private AND Father_educational_qualification = Secondary AND Mother_occupation = Private AND Mother_educational_qualification = Secondary AND Good_Class_Peers = Good AND Family_size = Big AND Child_position = Late THEN Ready	The readiness of the child is based here on parental secondary education, parental private job level, good class peers, big family size, and child position is late respectively. Child's peers are not affecting the result.
3	Father_occupation = Private AND Father_educational_qualification = Secondary AND Mother_occupation = Private AND Mother_educational_qualification = Secondary AND Family size = Big AND Child_position = Top THEN Ready	The readiness of the child is based here on parental secondary education, parental private job level, big family size, and child position is 1st or second respectively.
4	Father_occupation = Private AND Father_educational_qualification = Secondary AND Mother_occupation = Private AND Mother_educational_qualification = 1st_degree THEN Ready	The readiness of the child is based here on parental secondary and first academic degree education, parental private job level
5	Father_occupation = Private AND Father_educational_qualification = Secondary AND Mother_occupation = Private AND Child_position = Top AND Family_size = Big THEN Ready	The readiness of the child is based here on parental secondary education, parental private job level, big family size, and child position is third or above respectively.
6	Father_occupation = Private AND Father_educational_qualification = Secondary AND Mother_occupation = Private AND Family_size = Small AND Child_position = Top THEN Ready	The readiness of the child is based here on father secondary education, parental private job level, small family size, and child position is 1st or second respectively.
7	Father_occupation = Private AND Father_educational_qualification = Secondary AND Mother_occupation = Government AND Good_Neighbourhood_Peers = Good AND Family_size = Big THEN Ready	The readiness of the child is based here on father secondary education, father private job level, mother government job level and big family size respectively
8	Father_occupation = Private AND Father_educational_qualification = 1st_degree THEN Ready	The readiness of the child is based here on father 1st degree education and father private job level.
9	Father_occupation = Private AND Father_educational_qualification = 2nd_degree AND Mother_educational_qualification = PhD THEN Ready	This rule is very interesting because it fit my own child case.
10	Father_occupation = Government AND Weak_Neighbourhood_Peers = Good AND Good_Class_Peers = Good THEN Ready	The readiness of the child is based here on father government job level, good weak neighbourhood peer and good class peer respectively.

TABLE 9
SOME OF THE BEST RULES OBTAINED WITH THE J48 ALGORITHM

	Rules—Generated	Rules—Interpretation
1	Father_occupation = Private Mother_educational_qualification = 1st-degree	The father has a private job and the mother with academic degree.
2	Father_occupation = Private Mother_educational_qualification = 1st-degree Weak_class_Peers <= 4	The father has a private job and the mother with academic degree with minimum weak class peers.
3	Father_occupation = Private Mother_educational_qualification = 1st-degree Weak_class_Peers <= 4 Child_position > 1	The father has a private job and the mother with academic degree with minimum weak class peers and more than one child in the family.
4	Father_occupation = Private Mother_educational_qualification = Primary	The father has a private job and the mother with primary educational level, this is very typical for the Arab community inside Israel.
5	Father_occupation = Private Mother_educational_qualification = Secondary Weak_Neighbourhood_Peers <= 3	The father has a private job and the mother with secondary educational level, and child has 1-3 weak neighbourhood peers.
6	Father_occupation = Private Mother_educational_qualification = Secondary Mother_occupation = UnEmployed	The father has a private job and the mother with secondary educational level, and the mother unemployed, identical to rule (4).
7	Father_occupation = Private Mother_educational_qualification = Secondary Mother_occupation = Private Good_Neighbourhood_Peers >1 Family_size <= 4	The father and mother have a private job and the mother with secondary educational level, and child has more than one good neighbourhood peers and a small family members.
8	Father_occupation = Private Mother_educational_qualification = Secondary Mother_occupation = Private Good_Neighbourhood_Peers >1 Child_position >2	The father and mother have a private job and the mother with secondary educational level, and child has more than one good neighbourhood peers and a middle child in the family.
9	Father_occupation = Private Mother_educational_qualification = Secondary Mother_occupation = Private Good_Neighbourhood_Peers >1 Child_position <=2 Good_Class_Peers >2	The father and mother have a private job and the mother with secondary educational level, and child has more than one good neighbourhood peers and a middle child in the family and more than one good class peers.

9.7 The ID3 algorithm

ID3 builds a decision tree from a fixed set of examples. The resulting tree is used to classify future samples. The example has several attributes and belongs to a class (like yes or no). The leaf nodes of the decision tree contain the class name whereas a non-leaf node is a decision node. The decision node is an attribute test with each branch (to another decision tree) being a possible value of the attribute. ID3 uses information gain to help it decide which attribute goes into a decision node. The advantage of learning a decision tree is that a program, rather than a knowledge engineer, elicits knowledge from an expert.

10 RESULTS AND DISCUSSION

In this paper 5 different data mining algorithms were provided (Apriori, *k*-means, EM, ID3 and J48) for association, clustering and classification to predict if the child is ready according to socio-economic factors: father's education, father's job, mother's education, mother's job, family size, child position as well as siblings and friends.

Predicting school readiness can be a difficult task not only because it is a multifactor problem (in which there are a lot of personal, family, social, and economic factors that can be influential) but also because the available data are normally imbalanced. To resolve these problems, use of different DM algorithms and approaches for predicting school readiness had been discussed. Several experiments had been carried out using real data from different preschool classes in 4 different preschool children in the Arab community in Israel. Different classification, clustering and association approaches were applied for predicting the readiness status or final child performance at the end of the preschool. Furthermore it was shown that some approaches such as selecting the best attributes, cost-sensitive classification, and data balancing can also be very useful for improving accuracy.

It is important to notice that gathering information and preprocessing data were two very important tasks in this work. In fact, the quality and the reliability of the used information directly affect the results obtained. However, this is an arduous task that involves a lot of time. Specifically, data from a paper and pencil survey had been picked out and data from three different sources was integrated to form the final dataset.

The criteria described below

In general, regarding the DM approaches used and the classification, clustering and association results obtained, the main conclusions are as follows:

1. Classification, clustering and association algorithms can be used successfully in order to predict a child readiness for school and, in particular, to model the difference between ready and not ready children.
2. The number of attributes were reduced from the 71 initially available attributes to the best 11 attributes, obtaining fewer rules and conditions without losing classification performance.
3. Two different ways to address the problem of imbalanced data classification by rebalancing the data and considering different classification costs were shown. In fact, rebalancing of the data has been able to improve the classification results obtained in TN rate, Accuracy, and Geometric Mean.

Regarding the specific knowledge extracted from the DM models obtained, the main conclusions are as follows:

1. White box classification algorithms obtain models that can explain their predictions at a higher level of abstraction by IF-THEN rules. In this case, induction rule algorithms produce IF-THEN rules directly, decision trees and ID3 can be easily transformed into IF-THEN rules. IF-THEN rules are one of the most popular forms of knowledge representation, due to their simplicity and comprehensibility. These types of rules are easily understood and interpreted by non-expert DM users, such as instructors, and can be directly applied in decision making process.
2. Concerning the specific factor or attributes related with child readiness, there are some specific values that appear most frequently in the classification models obtained. For example, the values of parents' occupation that appear most frequently in the obtained classification rules are the value "Private". Other factor frequently associated with parents' education are being over 12 years of education, i.e. "Secondary" and "1st_Grade", also the family size is up to 5 members (Including both parents), and a middle child position in the family is the dominant.
3. This study was focused solely on social-demographic attributes to confirm the conventional results obtained only through empirically-based research.
4. Results have found a relationship between SES and school readiness. Children in higher SES group were more likely to be ready for school more than children in lower SES group.
5. The results approved the hypothesis that children from small families (three siblings or less) are more ready for school than children from large families (four siblings or more).
6. The results supported parental education's role in school readiness and found that level of maternal education was strongly related to school readiness of the child, mothers who are educated might be better able to invest in stimulating learning materials and engage in educational activities that may in turn promote learning in their children.
7. It is known that peer interactions are related to children's adjustment to school, but in this case the preschool children are still related to parents, brothers and sisters, so neighbourhood peers are not affecting the readiness in this stage, when a preschooler plays with brothers and sisters, he/she will receive prosocial behaviours from brothers and sisters, and school readiness will shed new light on the links between social-emotional development and children's early school success.

Starting from the previous models (rules and decision trees) generated by the DM algorithms, a system to alert the teacher and their parents about children who are potentially at risk of unready can be implemented. As an example of possible action, once children were found at risk, it proposed that they would be assigned to training activities in order to provide them with both improvement and guidance for motivating and trying to prevent child unready.

Present study shows that school and neighbourhood peers of the child are not always affecting the readiness value of the child. The investigation shows that other factors, father occupation, mother academic level, family size and child's position in the family, have got significant influence over the child's performance.

11 CONCLUSIONS

In this paper, an adapted methodology was presented for the application of data mining techniques to 5 different socio-economic features, trying to discover relevant parameters affect the child readiness.

The results show that the use of methods and data mining techniques are useful for the discovery of knowledge from information available. Clustering tests provided us with relevant information about the attributes that define each group. The classification and association tests supplied information significant of the key attributes that provide information to the knowledge-based rules to be used by the teachers and school.

This study will help the teacher to improve the child's performance, to identify those children who needed special attention to reduce unready induction and take appropriate action at the right time. Also this study will help parents to be aware of individual factors that may cause the child not to do well at school. This is also helpful for school administrators to better plan for better school-friends environment.

This is very important information for teachers and parents to know so they can improve the readiness status for the child in these early stages.

Finally, as the next step in this research, the aim is to:

1. Carry out more experiments using more data from different preschools (public, private and missionary) to test whether the same performance results are obtained with different DM approaches.
2. To focus on the school and neighbourhood peers attributes influencing the school readiness.
3. To examine and test more new attributes like: Marital status of parents, Location of kindergarten, Residence location, Care giver and more.
4. As a future work, it would be interesting to conduct a similar analysis on data collected from other countries to see whether similar patterns and conclusions can be observed.

12 REFERENCES

[1] Agarwal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD International Conference on Management of Data, Washington, DC.*, pp.207-216.

[2] Aluja, T. (2001). La minería de datos, entre la estadística y la inteligencia artificial, *Quaderns d'Estadística Invest. Operat.*, vol. 25, no. 3, pp. 479–498.

[3] Araque, F., Roldán, C., & Salguero, A. (2009). Factors influencing university drop out rates, *Computers & Education.*, vol. 53, no. 3, pp. 563–574.

[4] Ayers, E., Nugent, R., & Dean, N. (2009). A pilot study on logic proof tutoring using hints generated from historical student data. Paper presented at the 2nd International Conference on *Educational Data Mining (EDM'09)*, Cordoba, Spain.

[5] Baker, R. (2010). Data Mining for Education. To appear in McGaw, B., Peterson, P., & Baker, E. (Eds.) *International Encyclopaedia of Education* (3rd edition). Oxford, UK: Elsevier.

[6] Barnes, T., Desmarais, M., Romero, C., & Ventura, S. (2009). Educational Data Mining 2009: *2nd International Conference on Educational Data Mining*, _Proceedings. Cordoba, Spain

[7] Barnes, T., Stamper, J., Lehman, L., & Croy, M. (2008). A pilot study on logic proof tutoring using hints generated from historical student data. Paper presented at the *1st International Conference on Educational Data Mining* (EDM'08), Montreal, Canada.

[8] Barros, R. P., & Mendonça, R. (1999). *Costs and Benefits of Preschool Education in Brazil*. Rio de Janeiro: Institute of Applied Economic Research.

[9] Berger, L., et al. (2008). First-year maternal employment and child outcomes: Differences across racial and ethnic groups, *Children and Youth Services Review,* Vol. 30, pp. 365-387.

[10] Blair, C., & Diamond, A. (2008). Biological processes in prevention intervention:The promotion of selfregulation as a means of preventing school failure. Development and Psychopathology, 20, 899- 911.

[11] Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children*, Vol. 7, No. 2, pp. 55-71.

[12] Brooks-Gunn, J., Han, W. J., & Waldfogel, J. (2002). Maternal employment and child cognitive outcomes in the first three years of life: The NICHD study of early child care. *Child Development*, Vol. 73, No. 4, pp. 1052-1072.

[13] Brusilovsky, P., & Peylo, C. (2003). Adaptive and Intelligent Web-based Educational Systems // *International Journal of Artificial Intelligence in Education*. – Vol. 13, pp. 156-169.

[14] Brusilovsky, P., & Miller, P. (2001) 'Course Delivery Systems for the Virtual University'. In: Della Senta, T., Tschang, T. (eds.): Access to Knowledge: *New Information Technologies and the Emergence of the Virtual University*, pp. 167-206. Elsevier Science, Amsterdam.

[15] Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User Adapted Interaction,* Vol. 11, No. 1/2, pp. 87-110, Ten Year Anniversary Issue (Alfred Kobsa, ed.).

[16] Calders, T., & Pechenizkiy, M. (2012). Cost-Sensitive Classification Problem. In: Workshop on Techning Machine Learning (TML) at ICML.

[17] Castro, F., Vellido, A., Nebot, A., & Mugica, F. (2007). Applying Data Mining Techniques to e-Learning Problems. In: Jain, L.C., Tedman, R. and Tedman, D. (eds.) Evolution of Teaching and Learning Paradigms in Intelligent Environment. Studies in Computational Intelligence, 62, Springer-Verlag,, pp. 183-221.

[18] Coley, R. J. (2002). *An uneven start: Indicators of inequality in school readiness*, Available at: http://www.ets.org/Media/Research/pdf/PICUNEVENSTART.pdf (Accessed: April 5, 2009).

[19] Connell, C. M. (2001). The relationship between paren-

tal behaviors, academic readiness and social skills development in at-risk kindergarten students. *Dissertation Abstracts International*, Vol. 62, No. 2, 1072B. (UMI No. 3006016)

[20] Currie, J., & Thomas, D. (1996). Does Head Start help Hispanic children? Labor and population program, working paper series 96-17 (Report No. DRU-1528-RC). Bethesda, MD: National Institute of Child Health and Human Development. (ERIC Document Reproduction Service No. 404008).

[21] Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., & Klebanov, P. (2007). School readiness and later achievement, *Developmental Psychology*, Vol. 43, No. 6, pp. 1428-1446.

[22] Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children*, Vol. 15, No. 1, pp. 35-54.

[23] Farkas, G., & Hibel, J. (2008). Being unready for school: Factors affecting risk and resilience In A. Booth & A. Crouter (Eds.), Disparities in School Readiness: How families contribute to transitions into school (pp. 3-28). New York: Lawrence Erlbaum Associates.

[24] Farran, D. C. (2000). Another decade of intervention for children who are low income or disabled: what do we know now? In J. Shonkoff & S. Meisels (Eds.), Handbook of early intervention, second edition (pp. 510–548). Cambridge, England: Cambridge University Press.

[25] Farran, D. C., & Son-Yarbroug, W. (2001). Title I funded preschools as a developmental context for children's play and verbal behaviors. *Early Childhood Research Quarterly*, Vol. 16 , pp. 245–262.

[26] Farver, J. A. M., Xu, Y., Eppe, S., & Lonigan, C. J. (2006). Home environments and young Latino children's school readiness. *Early Childhood Research Quarterly*, Vol. 21, No. 2, pp. 196-212.

[27] Fayyad, U., Piatetsky-shapiro G., & Smyth. P. (1996). From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence*, Vol. 17, pp. 37-54.

[28] Fisher, D., Smith, M., & Welser, H. (2006). You are who you talk to: Detecting roles in usenet newsgroups. In Proceedings of the 39th Annual HICSS. IEEE Computer Society.

[29] Fowler, M., & Cross, A. (1986). Preschool risk factors as predictors of early school performance. Journal of Developmental and Behavioral Pediatrics, Vol. 7, No. 4, pp. 237-241.

[30] García, E., Romero, C., Ventura, S., & Castro, C. (2009). An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering. *Journal for User Model and User Adapted Interaction*, Vol. 19, No. 1-2, pp. 99-132.

[31] Grantham-McGregor. S., et al. (2007). 'Developmental Potential in the First 5 Years for Children in Developing Countries', *The Lancet*, Vol. 369, No. 9555, 6–12 January 2007, pp. 60–70.

[32] Guo, Q., & Zhang, M. (2009). Implement web learning environment based on data mining. *Elsevier Knowledge Based Systems*, Vol. 22, No. 6, pp. 439-442.

[33] Han, J., & Kamber, M,. & Pei, J. (2011). *Data Mining Concepts and Techniques*, 3rd Edition, Morgan Kaufmann.

[34] Han, J. & Kamber, M. (2001), Data Mining: Concepts and Techniques, Morgan Kaufmann.

[35] Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Brookes Publishing.

[36] Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. In WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web Mining and Social Network Analysis, pp. 56–66.

[37] Kamerman, S. B. (2002) *Early Childhood Care and Education and Other Family Policies and Programs in South-East Asia*, Paris: United Nations Educational, Scientific and Cultural Organization.

[38] Koedinger, K., Cunningham, K., Skogsholm, A., & Leber, B. (2008). An open repository and analysis tools for fine-grained, longitudinal learner data. In *1st International Conference on Educational Data Mining*, Montreal, pp. 157-166.

[39] Kotsiantis, S. (2009). "Educational data mining: A case study for predicting dropout—prone students," *Int. J. Know. Eng. Soft Data Paradigms*, Vol. 1, No. 2, pp. 101–111.

[40] Kotsiantis, S., Patriarcheas, K., & Xenos, M. (2010). "A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education", *Knowledge. Based Syst*., Vol. 23, No. 6, pp. 529–535.

[41] Krebs, V. (2002). Mapping Networks of Terrorist Cells. *Connections*. Vol. 24, No. 3, pp. 43-52.

[42] Ladd, G. W. (1990). Having friends, Keeping friends, making friends, and being liked by peers in the classroom: Predictors of children's early school adjustment? *Child Development*, Vol. 61, pp. 1081-1100.

[43] Levine, M. (1966). Residential change and school adjustment. *Mental Health Journal*, Vol. 2, pp. 61-69.

[44] Linoff, M. J. A. B. G. (2004), Data Mining Techniques: for marketing, sales, and customer relationship management, Wiley Publishing.

[45] Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). "Dropout prediction in e-learning courses through the combination of machine learning techniques," Comput. Educ., Vol. 53, No. 3, pp. 950–965.

[46] Madhyastha, T., & Hunt, E. (2009). Mining diagnostic assessment data for concept similarity. *Journal of Educational Data Mining*, Vol. 1, No. 1, pp. 1-19.

[47] Maimon, O., & Rokach, L (2010) *Data Mining and Knowledge Discovery Handbook*, Second Edition, Springer: New York .

[48] Marks, A. K., & Coll, C. G. (2007). Psychological and

demographic correlates of early academic skill development among American Indian and Alaska Native Youth: A growth modeling study. *Developmental Psychology*, Vol. 43, No. 3, pp. 663-674.

[49] Márquez-Vera, C., Morales, C. R., & Soto, S. V. (2013). Predicting School Failure and Dropout by Using Data Mining Techniques. *IEEE Journal of Latin-American Learning Technologies*, Vol. 8, No. 1.

[50] Márquez-Vera, C., Romero, C., & Ventura, S. (2011). Predicting School Failure Using Data Mining. *Expert Systems With Applications*, Vol. 38, No. 12, pp. 15020-15031.

[51] McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, Vol. 53, No. 2, pp. 185-204.

[52] McWayne, C. M., Fantuzzo, J. W., & McDermott, P. A. (2004). Preschool competency in context: An investigation of the unique contribution of child competencies to early academic success. *Developmental Psychology*, Vol. 40, pp. 633-645.

[53] Mostow, J., & Beck., J. (2006). Some useful tactics to modify, map and mine data from intelligent tutors. In *Journal Natural Language Engineering*, Vol. 12 , No. 2, pp. 195-208.

[54] Nagarajan, P., & Wiselin Jiji, G. (2010) 'Online Educational System (e- learning) ', *International Journal of u- International Journal of u- and e- Service, Science and Technology*, Vol. 3(No. 4), pp. 37-48.

[55] National Center for Education Statistics (2001). U.S. Department of Education, ECLS-K Base Year Data Files and Electronic Codebook. Retrieved June 4, 2008, from http://nces.ed.gov/ecls/kindergarten.asp

[56] Parker, A. (1999). "A study of variables that predict dropout from distance education," *International Journal Education Technology*, Vol. 1, No. 2, pp. 1–11.

[57] Pavlik, P., Cen, H., & Koedinger, K. (2009). Learning factors transfer analysis: Using learning curve analysis to automatically generate domain models. Paper presented at the *2nd International Conference on Educational Data Mining* (EDM'09), Cordoba, Spain.

[58] Pechenizkiy, M., Trčka, N., Vasilyeva, E., Aalst, W., & De Bra, P. (2009). Process mining online assessment data. Paper presented at the *2nd International Conference on Educational Data Mining* (EDM'09), Cordoba, Spain.

[59] Perez, S. M., & Martinez, D. (1993). *State of Hispanic America 1993: Toward a Latino anti-poverty agenda*. (UD No. 029424). Washington DC: National Council of La Raza. (ERIC Document Reproduction Service No. ED360465)

[60] Praag, C. (2002). *The Social and Cultural Report 2002*. The Netherlands Institute for Social Research.

[61] Pyle, D. (1999). Data Preparation for Data Mining. Morgan Kaufmann Publishers, Los Altos, California.

[62] Quadril, M. N., & Kalyankar, N. V. (2010). "Drop out feature of student data for academic performance using decision tree techniques", *Global Journal. of Computer Science and Technology*, vol. 10, pp. 2–5.

[63] Quinlan, J. R., (1986). Induction to decision trees. *Machine Learning*, Vol. 1, No, 1, pp. 81–106.

[64] Rabbany, R., Takaffoli, M., & Zaiane, O. (2011). Analyzing Participation of Students in Online Courses Using Social Network Analysis Techniques. In: Proceedings of the *4th International Conference on Educational Data Mining*. pp. 21-30.

[65] Rock, D. A., & Stenner, A. J. (2005). Assessment issues in the testing of children at school entry. *The Future of Children*, Vol. 15, No. 1, pp. 15-34.

[66] Rodriguez, E. T. (2008). The home literacy environment: Predictors of trajectories across the first five years and relations to children's school readiness at prekindergarten. *Dissertation Abstracts International*, Vol. 69, No. 3, 1988B. (UMI No. 3308289)

[67] Roe, K. V., & Bronstein, R. (1988). Maternal education and cognitive processing at three months as shown by the infants' vocal response to mother vs. stranger. *International Journal of Behavioral Development*, Vol. 11, No. 3, pp. 389-393.

[68] Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of *the Art. IEEE Transactions on Systems, Man, and Cybernetics*, Part C: Applications and Reviews, Vol. 40, No. 6, pp. 601-618.

[69] Romero, C., & Ventura, S. (2007). "Educational data mining: A survey from 1995 to 2005", *Expert Systems with Applications*, Vol. 33, No. 1, pp. 135–146.

[70] Romero, C., Ventura, S., Zafra, A., & de Bra, P. (2009). Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems. Paper presented at *Computers & Education*, Vol. 53, pp. 828–840.

[71] Romero, C., Ventura, S., Espejo, P., & Hervas, C. (2008). Data mining algorithms to classify students. Paper presented at *1st International Conference on Educational Data Mining* (EDM'08), Montreal, Canada.

[72] Rouse, C., Brooks-Gunn, J., & McLanahan, S. (2005). Introducing the Issue. *The Future of Children*, Vol. 15, No. 1, pp. 5-14.

[73] Sameroff, A. J. (1998). Management of clinical problems and emotional care: environmental risk factors in infancy. *Pediatrics*, Vol. 102, No. 5, pp. 1287-1292.

[74] Scott, R., & Seifert, K. (1975). Family size and learning readiness profiles of socioeconomically disadvantaged preschool whites. *Journal of Psychology: Interdisciplinary and Applied*, Vol. 89, No. 1, pp. 3-7.

[75] Smyth, D. H. . H. M. . P. (2001), Principles of Data Mining, MIT.

[76] UNESCO (2007) *EFA Global Monitoring Report 2007: Strong foundations – Early childhood care and education*, Paris: UNESCO.

[77] Wachs, T. D., & Camli, O. (1991). Do ecological or individual characteristics mediate the influence of the physical environment upon maternal behavior? *Developmental Psychology*, Vol. 11, No. 3, pp. 249-264.

[78] Willms, J. D. (2002). *Standards of care: Investments to im-*

prove children's educational outcomes in Latin America. In M. E.Young (Ed.), From early child development to human development (pp. 81–122). Washington, DC: The World Bank.

[79] Willms, J. D., & Somers, M. A. (2001). Family, classroom and school effects on children's educational outcomes in Latin America. *International Journal of School Effectiveness and Improvement*, Vol. 12, No. 4, pp. 409-445.

[80] Witten, H., & Frank, E. (2011). Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, 3rd edition, Morgan Kaufmann.

[81] Ye, P., & Baldwin, T. (2005). Semantic role labeling of prepositional phrases. In The Second *International Joint Conference on Natural Language Processing*, Jeju Island, Korea, pp. 779–791.

[82] Zill, N., & et al. (1995). School readiness and children's developmental status. Urbana, IL: ERIC Clearinghouse on Elementary and Early Childhood Education. Report: EDO PS. (ERIC Document Reproduction Service Report No. ED389475).

[83] MacQueen, J. (1967) Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 1, 281-297.

[84] Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society Series B (Statistical Methodology) 39(1):1-38.

APPENDIX: WEKA DATA MINING RESULTS

The APriori algorithm WEKA results

```
=== Run information ===
Scheme:       weka.associations.Apriori -N 100 -T 0 -C 0.9 -D 0.05 -U 1.0 -
M 0.3 -S -1.0 -c -1
Relation:     Ready2Learn-weka.filters.unsupervised.attribute.Remove-R1-
2,9-11,16-24
Instances:    306
Attributes:   11
              Mother_educational_qualification
              Father_educational_qualification
              Mother_occupation
              Father_occupation
              Family_size
              Child_position
              Good_Class_Peers
              Weak_Class_Peers
              Good_Neighbourhood_Peers
              Weak_Neighbourhood_Peers
              Ready4School
=== Associator model (full training set) ===

Apriori
=======
Minimum support: 0.4 (122 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 12

Generated sets of large itemsets:
Size of set of large itemsets L(1): 14
Size of set of large itemsets L(2): 48
Size of set of large itemsets L(3): 60
Size of set of large itemsets L(4): 28
Size of set of large itemsets L(5): 5

Best rules found:
 1. Ready4School=1 273 ==> Father_occupation=1 273    <conf:(1)>
  lift:(1) lev:(0) [0] conv:(0.89)
 2. Mother_occupation=1 Ready4School=1 244 ==> Father_occupation=1
  244    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.8)
 3. Child_position=1 Ready4School=1 223 ==> Father_occupation=1 223
  <conf:(1)> lift:(1) lev:(0) [0] conv:(0.73)
 4. Mother_occupation=1 Child_position=1 Ready4School=1 199 ==>
  Father_occupation=1 199    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.65)
 5. Father_educational_qualification=0 Ready4School=1 193 ==> Fa-
  ther_occupation=1 193    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.63)
 6. Good_Class_Peers=0 189 ==> Father_occupation=1 189    <conf:(1)>
  lift:(1) lev:(0) [0] conv:(0.62)
 7. Family_size=0 Ready4School=1 178 ==> Father_occupation=1 178
  <conf:(1)> lift:(1) lev:(0) [0] conv:(0.58)
 8. Mother_educational_qualification=0 Ready4School=1 177 ==> Fa-
  ther_occupation=1 177    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.58)
 9. Good_Class_Peers=0 Ready4School=1 172 ==> Fa-
  ther_occupation=1 172    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.56)
10. Mother_occupation=1 Good_Class_Peers=0 169 ==> Fa-
  ther_occupation=1 169    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.55)
11. Father_educational_qualification=0 Mother_occupation=1
  Ready4School=1 168 ==> Father_occupation=1 168    <conf:(1)> lift:(1)
  lev:(0) [0] conv:(0.55)
12. Weak_Neighbourhood_Peers=0 165 ==> Father_occupation=1 165
  <conf:(1)> lift:(1) lev:(0) [0] conv:(0.54)
13. Mother_educational_qualification=0 Fa-
  ther_educational_qualification=0 Ready4School=1 164 ==> Fa-
  ther_occupation=1 164    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.54)
14. Good_Neighbourhood_Peers=0 Ready4School=1 157 ==> Fa-
  ther_occupation=1 157    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.51)
15. Mother_occupation=1 Family_size=0 Ready4School=1 156 ==> Fa-
  ther_occupation=1 156    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.51)
16. Father_educational_qualification=0 Child_position=1
  Ready4School=1 155 ==> Father_occupation=1 155    <conf:(1)> lift:(1)
  lev:(0) [0] conv:(0.51)
17. Mother_educational_qualification=0 Mother_occupation=1
  Ready4School=1 154 ==> Father_occupation=1 154    <conf:(1)> lift:(1)
  lev:(0) [0] conv:(0.5)
18. Mother_occupation=1 Good_Class_Peers=0 Ready4School=1 154
  ==> Father_occupation=1 154    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.5)
19. Child_position=1 Good_Class_Peers=0 151 ==> Fa-
  ther_occupation=1 151    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.49)
20. Mother_occupation=1 Weak_Neighbourhood_Peers=0 148 ==> Fa-
  ther_occupation=1 148    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.48)
21. Weak_Neighbourhood_Peers=0 Ready4School=1 148 ==> Fa-
  ther_occupation=1 148    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.48)
22. Weak_Class_Peers=0 Ready4School=1 143 ==> Fa-
  ther_occupation=1 143    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.47)
23. Weak_Class_Peers=1 142 ==> Father_occupation=1 142
  <conf:(1)> lift:(1) lev:(0) [0] conv:(0.46)
24. Mother_educational_qualification=0 Fa-
  ther_educational_qualification=0 Mother_occupation=1
  Ready4School=1 141 ==> Father_occupation=1 141    <conf:(1)> lift:(1)
  lev:(0) [0] conv:(0.46)
25. Mother_educational_qualification=0 Child_position=1
  Ready4School=1 140 ==> Father_occupation=1 140    <conf:(1)> lift:(1)
  lev:(0) [0] conv:(0.46)
26. Father_educational_qualification=0 Good_Class_Peers=0 138 ==>
  Father_occupation=1 138    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.45)
27. Mother_occupation=1 Good_Neighbourhood_Peers=0
  Ready4School=1 138 ==> Father_occupation=1 138    <conf:(1)> lift:(1)
  lev:(0) [0] conv:(0.45)
28. Child_position=1 Good_Class_Peers=0 Ready4School=1 136 ==>
  Father_occupation=1 136    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.44)
29. Child_position=1 Weak_Neighbourhood_Peers=0 135 ==> Fa-
  ther_occupation=1 135    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.44)
30. Father_educational_qualification=0 Mother_occupation=1
  Child_position=1 Ready4School=1 135 ==> Father_occupation=1 135
  <conf:(1)> lift:(1) lev:(0) [0] conv:(0.44)
31. Mother_occupation=1 Child_position=1 Good_Class_Peers=0 134
  ==> Father_occupation=1 134    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.44)
32. Child_position=1 Good_Neighbourhood_Peers=0 Ready4School=1
  134 ==> Father_occupation=1 134    <conf:(1)> lift:(1) lev:(0) [0]
  conv:(0.44)
33. Mother_occupation=1 Weak_Neighbourhood_Peers=0
  Ready4School=1 133 ==> Father_occupation=1 133    <conf:(1)> lift:(1)
  lev:(0) [0] conv:(0.43)
34. Good_Neighbourhood_Peers=1 132 ==> Father_occupation=1 132
  <conf:(1)> lift:(1) lev:(0) [0] conv:(0.43)
35. Weak_Class_Peers=1 Ready4School=1 130 ==> Fa-
  ther_occupation=1 130    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.42)
36. Mother_occupation=1 Weak_Class_Peers=0 Ready4School=1 129
  ==> Father_occupation=1 129    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.42)
37. Mother_educational_qualification=0 Fa-
  ther_educational_qualification=0 Child_position=1 Ready4School=1
  129 ==> Father_occupation=1 129    <conf:(1)> lift:(1) lev:(0) [0]
  conv:(0.42)
38. Family_size=0 Child_position=1 Ready4School=1 128 ==> Fa-
  ther_occupation=1 128    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.42)
39. Family_size=0 Good_Class_Peers=0 127 ==> Father_occupation=1
  127    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.42)
40. Mother_occupation=1 Weak_Class_Peers=1 126 ==> Fa-
  ther_occupation=1 126    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.41)
41. Weak_Neighbourhood_Peers=1 Ready4School=1 125 ==> Fa-
  ther_occupation=1 125    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.41)
42. Father_educational_qualification=0 Family_size=0 Ready4School=1
  124 ==> Father_occupation=1 124    <conf:(1)> lift:(1) lev:(0) [0]
  conv:(0.41)
43. Father_educational_qualification=0 Good_Class_Peers=0
  Ready4School=1 123 ==> Father_occupation=1 123    <conf:(1)> lift:(1)
  lev:(0) [0] conv:(0.4)
44. Mother_educational_qualification=0 Good_Class_Peers=0 122 ==>
  Father_occupation=1 122    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.4)
45. Mother_educational_qualification=0 Mother_occupation=1
  Child_position=1 Ready4School=1 122 ==> Father_occupation=1 122
  <conf:(1)> lift:(1) lev:(0) [0] conv:(0.4)
```

46. Mother_occupation=1 273 ==> Father_occupation=1 272 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.45)
47. Child_position=1 253 ==> Father_occupation=1 252 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.41)
48. Mother_occupation=1 Child_position=1 225 ==> Father_occupation=1 224 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.37)
49. Father_educational_qualification=0 218 ==> Father_occupation=1 217 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.36)
50. Mother_educational_qualification=0 199 ==> Father_occupation=1 198 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.33)
51. Family_size=0 197 ==> Father_occupation=1 196 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.32)
52. Father_educational_qualification=0 Mother_occupation=1 191 ==> Father_occupation=1 190 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.31)
53. Mother_educational_qualification=0 Father_educational_qualification=0 184 ==> Father_occupation=1 183 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.3)
54. Father_educational_qualification=0 Child_position=1 177 ==> Father_occupation=1 176 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.29)
55. Mother_educational_qualification=0 Mother_occupation=1 175 ==> Father_occupation=1 174 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.29)
56. Good_Neighbourhood_Peers=0 174 ==> Father_occupation=1 173 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.28)
57. Mother_occupation=1 Family_size=0 173 ==> Father_occupation=1 172 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.28)
58. Weak_Class_Peers=0 164 ==> Father_occupation=1 163 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.27)
59. Mother_educational_qualification=0 Father_educational_qualification=0 Mother_occupation=1 160 ==> Father_occupation=1 159 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.26)
60. Mother_educational_qualification=0 Child_position=1 159 ==> Father_occupation=1 158 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.26)
61. Father_educational_qualification=0 Mother_occupation=1 Child_position=1 155 ==> Father_occupation=1 154 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.25)
62. Mother_occupation=1 Good_Neighbourhood_Peers=0 153 ==> Father_occupation=1 152 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.25)
63. Child_position=1 Good_Neighbourhood_Peers=0 149 ==> Father_occupation=1 148 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.24)
64. Mother_occupation=1 Weak_Class_Peers=0 147 ==> Father_occupation=1 146 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.24)
65. Mother_educational_qualification=0 Father_educational_qualification=0 Child_position=1 146 ==> Father_occupation=1 145 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.24)
66. Family_size=0 Child_position=1 144 ==> Father_occupation=1 143 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.24)
67. Weak_Neighbourhood_Peers=1 141 ==> Father_occupation=1 140 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.23)
68. Mother_educational_qualification=0 Mother_occupation=1 Child_position=1 140 ==> Father_occupation=1 139 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.23)
69. Father_educational_qualification=0 Family_size=0 139 ==> Father_occupation=1 138 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.23)
70. Child_position=1 Weak_Class_Peers=0 137 ==> Father_occupation=1 136 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.22)
71. Mother_educational_qualification=0 Family_size=0 129 ==> Father_occupation=1 128 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.21)
72. Mother_occupation=1 Child_position=1 Good_Neighbourhood_Peers=0 129 ==> Father_occupation=1 128 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.21)
73. Mother_educational_qualification=0 Father_educational_qualification=0 Mother_occupation=1 Child_position=1 127 ==> Father_occupation=1 126 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.21)
74. Mother_occupation=1 Weak_Neighbourhood_Peers=1 125 ==> Father_occupation=1 124 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.2)
75. Mother_occupation=1 Family_size=0 Child_position=1 125 ==> Father_occupation=1 124 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.2)
76. Father_educational_qualification=0 Good_Neighbourhood_Peers=0 123 ==> Father_occupation=1 122 <conf:(0.99)> lift:(1) lev:(0) [0] conv:(0.2)
77. Mother_educational_qualification=0 Family_size=0 129 ==> Father_educational_qualification=0 122 <conf:(0.95)> lift:(1.33) lev:(0.1) [30] conv:(4.64)
78. Mother_educational_qualification=0 Ready4School=1 177 ==> Father_educational_qualification=0 164 <conf:(0.93)> lift:(1.3) lev:(0.12) [37] conv:(3.64)
79. Mother_educational_qualification=0 Father_occupation=1 Ready4School=1 177 ==> Father_educational_qualification=0 164 <conf:(0.93)> lift:(1.3) lev:(0.12) [37] conv:(3.64)
80. Mother_educational_qualification=0 Ready4School=1 177 ==> Father_educational_qualification=0 Father_occupation=1 164 <conf:(0.93)> lift:(1.31) lev:(0.13) [38] conv:(3.68)
81. Mother_educational_qualification=0 199 ==> Father_educational_qualification=0 184 <conf:(0.92)> lift:(1.3) lev:(0.14) [42] conv:(3.58)
82. Mother_educational_qualification=0 Father_occupation=1 198 ==> Father_educational_qualification=0 183 <conf:(0.92)> lift:(1.3) lev:(0.14) [41] conv:(3.56)
83. Mother_educational_qualification=0 Child_position=1 Ready4School=1 140 ==> Father_educational_qualification=0 129 <conf:(0.92)> lift:(1.29) lev:(0.1) [29] conv:(3.36)
84. Mother_educational_qualification=0 Father_occupation=1 Child_position=1 Ready4School=1 140 ==> Father_educational_qualification=0 129 <conf:(0.92)> lift:(1.29) lev:(0.1) [29] conv:(3.36)
85. Mother_educational_qualification=0 Child_position=1 Ready4School=1 140 ==> Father_educational_qualification=0 Father_occupation=1 129 <conf:(0.92)> lift:(1.3) lev:(0.1) [29] conv:(3.39)
86. Mother_educational_qualification=0 199 ==> Father_educational_qualification=0 Father_occupation=1 183 <conf:(0.92)> lift:(1.3) lev:(0.14) [41] conv:(3.4)
87. Mother_educational_qualification=0 Child_position=1 159 ==> Father_educational_qualification=0 146 <conf:(0.92)> lift:(1.29) lev:(0.11) [32] conv:(3.27)
88. Mother_educational_qualification=0 Father_occupation=1 Child_position=1 158 ==> Father_educational_qualification=0 145 <conf:(0.92)> lift:(1.29) lev:(0.11) [32] conv:(3.25)
89. Mother_educational_qualification=0 Mother_occupation=1 Ready4School=1 154 ==> Father_educational_qualification=0 141 <conf:(0.92)> lift:(1.29) lev:(0.1) [31] conv:(3.16)
90. Mother_educational_qualification=0 Mother_occupation=1 Father_occupation=1 Ready4School=1 154 ==> Father_educational_qualification=0 141 <conf:(0.92)> lift:(1.29) lev:(0.1) [31] conv:(3.16)
91. Mother_educational_qualification=0 Mother_occupation=1 Ready4School=1 154 ==> Father_educational_qualification=0 Father_occupation=1 141 <conf:(0.92)> lift:(1.29) lev:(0.1) [31] conv:(3.2)
92. Weak_Class_Peers=1 142 ==> Ready4School=1 130 <conf:(0.92)> lift:(1.03) lev:(0.01) [3] conv:(1.18)
93. Father_occupation=1 Weak_Class_Peers=1 142 ==> Ready4School=1 130 <conf:(0.92)> lift:(1.03) lev:(0.01) [3] conv:(1.18)
94. Weak_Class_Peers=1 142 ==> Father_occupation=1 Ready4School=1 130 <conf:(0.92)> lift:(1.03) lev:(0.01) [3] conv:(1.18)
95. Mother_educational_qualification=0 Mother_occupation=1 175 ==> Father_educational_qualification=0 160 <conf:(0.91)> lift:(1.28) lev:(0.12) [35] conv:(3.15)
96. Mother_educational_qualification=0 Mother_occupation=1 Father_occupation=1 174 ==> Father_educational_qualification=0 159 <conf:(0.91)> lift:(1.28) lev:(0.11) [35] conv:(3.13)
97. Mother_educational_qualification=0 Child_position=1 159 ==> Father_educational_qualification=0 Father_occupation=1 145 <conf:(0.91)> lift:(1.29) lev:(0.11) [32] conv:(3.08)
98. Mother_occupation=1 Good_Class_Peers=0 169 ==> Ready4School=1 154 <conf:(0.91)> lift:(1.02) lev:(0.01) [3] conv:(1.14)
99. Mother_occupation=1 Father_occupation=1 Good_Class_Peers=0 169 ==> Ready4School=1 154 <conf:(0.91)> lift:(1.02) lev:(0.01) [3] conv:(1.14)
100. Mother_occupation=1 Good_Class_Peers=0 169 ==> Father_occupation=1 Ready4School=1 154 <conf:(0.91)> lift:(1.02) lev:(0.01) [3] conv:(1.14)

The K-Means Algorithm WEKA Results

=== Run information ===
Scheme: weka.clusterers.SimpleKMeans -N 5 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10
Relation: Ready2Learn-weka.filters.unsupervised.attribute.Remove-R1-2-weka.filters.unsupervised.attribute.Remove-R7-9-weka.filters.unsupervised.attribute.Remove-R11-19
Instances: 306
Attributes: 11
Mother_educational_qualification
Father_educational_qualification
Mother_occupation
Father_occupation
Family_size
Child_position
Good_Class_Peers
Weak_Class_Peers
Good_Neighbourhood_Peers
Weak_Neighbourhood_Peers
Ready4School
Test mode: split 66% train, remainder test

kMeans
=======
Number of iterations: 9
Within cluster sum of squared errors: 320.5208760921712
Missing values globally replaced with mean/mode

Cluster centroids:

		Cluster#				
Attribute	Full Data	0	1	2	3	4
	(306)	(84)	(73)	(87)	(31)	(31)
Mother_educational_qualification	H.School	H.School	1st_degree	H.School	1st_degree	Secondary
Father_educational_qualification	H.School	H.School	1st_degree	H.School	H.School	Secondary
Mother_occupation	Private	Private	Private	Private	Private	Government
Father_occupation	Private	Private	Private	Private	Private	Private
Family_size	4.7026	4.9643	4.6849	4.5862	4.3548	4.7097
Child_position	1.9346	2.2738	1.7671	1.8621	1.5806	1.9677
Good_Class_Peers	2.9641	4.0476	2.9726	1.9195	3.2903	2.6129
Weak_Class_Peers	2.8203	3.0357	2.9041	2.5517	3.1290	2.4839
Good_Neighbourhood_Peers	2.8725	2.7976	2.8630	3.0920	2.7097	2.6452
Weak_Neighbourhood_Peers	2.8268	3.2857	2.8082	2.1954	2.8065	3.4194
Ready4School	Yes	Yes	Yes	Yes	Yes	Yes

Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on test split ===

kMeans
======
Number of iterations: 14
Within cluster sum of squared errors: 228.14912670587015
Missing values globally replaced with mean/mode

Cluster centroids:

		Cluster#				
Attribute	Full Data	0	1	2	3	4
	(201)	(39)	(35)	(38)	(37)	(52)
Mother_educational_qualification	H.School	H.School	H.School	H.School	H.School	1st_degree
Father_educational_qualification	H.School	H.School	H.School	H.School	H.School	1st_degree
Mother_occupation	Private	Private	Private	Private	Private	Private
Father_occupation	Private	Private	Private	Private	Private	Private
Family_size	4.7164	4.6667	4.5429	4.8158	4.7297	4.7885
Child_position	1.9353	1.7949	1.9429	2.1316	2	1.8462
Good_Class_Peers	3.0149	2.7692	3.1143	2.6053	3.6757	2.9615
Weak_Class_Peers	2.8358	2.9487	1.9143	1.9474	4.1892	3.0577
Good_Neighbourhood_Peers	2.9701	4.2821	3.9429	1.7895	1.7568	3.0577
Weak_Neighbourhood_Peers	2.7015	1.9231	4.2286	1.5789	3.2973	2.6538
Ready4School	Yes	Yes	Yes	Yes	Yes	Yes

Time taken to build model (percentage split) : 0.02 seconds

Clustered Instances
0 15 (14%)
1 23 (22%)
2 17 (16%)
3 23 (22%)
4 27 (26%)

The Expectation-Maximisation (Em) Algorithm Weka Results

=== Run information ===
Scheme: weka.clusterers.EM -I 100 -N 4 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -num-slots 1 -S 200
Relation: Ready2Learn-weka.filters.unsupervised.attribute.Remove-R1-2,9-11,16-24
Instances: 306
Attributes: 11
Mother_educational_qualification
Father_educational_qualification
Mother_occupation
Father_occupation
Family_size
Child_position
Good_Class_Peers
Weak_Class_Peers
Good_Neighbourhood_Peers
Weak_Neighbourhood_Peers
Ready4School
Test mode: split 66% train, remainder test

=== Clustering model (full training set) ===

EM
==
Number of clusters: 4
Number of iterations performed: 26

	Cluster			
	0	1	2	3
	(0.28)	(0.55)	(0.03)	(0.14)
Mother_educational_qualification				
Primary	1.0814	1.0554	2.9917	4.8716
Secondary	3.2824	153.2146	2.4290	38.0740
1st_degree	79.6616	15.3416	6.9338	3.0630
2nd_degree	3.4803	1.3248	1.0118	1.1831
MD	2.4575	1.0126	1.5276	1.0023
PhD	1.0180	1.0016	1.0003	1.9801
[total]	90.9811	172.9506	15.8943	50.1740
Father_educational_qualification				
Primary	1.0821	1.0583	2.9918	5.8678
Secondary	16.9678	154.3899	6.3541	37.2882
1st_degree	65.7893	14.1211	3.2292	2.8604
2nd_degree	1.5708	1.2655	1.0100	2.1537
MD	4.5712	1.1158	1.3092	1.0038
PhD	1.0000	1.0000	1.0000	1.0000
[total]	90.9811	172.9506	15.8943	50.1740
Mother_occupation				
UnEmployed	6.8854	20.8486	2.7372	6.5288
Private	51.9359	136.8048	1.5246	34.7347
Government	29.1598	12.2972	8.6324	5.9106
[total]	87.9811	169.9506	12.8943	47.174
Father_occupation				
UnEmployed	1.0022	1.9410	1.0030	1.0538
Private	75.8401	153.3461	9.9799	40.8338
Government	11.1388	14.6634	1.9113	5.2864
[total]	87.9811	169.9506	12.8943	47.1740
Family_size				
Big	60.3667	94.1779	1.3863	45.0691

Small	26.6144	74.7727	10.5080	1.1049
[total]	86.9811	168.9506	11.8943	46.174
Child_position				
Late	11.5188	4.1964	1.0385	40.2463
Top	75.4623	164.7542	10.8557	5.9277
[total]	86.9811	168.9506	11.8943	46.1740
Good_Class_Peers				
Good	34.1558	67.3908	5.5991	13.8542
Weak	52.8254	101.5597	6.2951	32.3198
[total]	86.9811	168.9506	11.8943	46.174
Weak_Class_Peers				
Good	36.6143	85.5786	1.2019	22.6051
Weak	50.3668	83.3719	10.6924	23.5689
[total]	86.9811	168.9506	11.8943	46.174
Good_ Neighbourhood _Peers				
Good	38.6347	67.2160	8.3640	21.7853
Weak	48.3464	101.7345	3.5303	24.3887
[total]	86.9811	168.9506	11.8943	46.1740
Weak_ Neighbourhood _Peers				
Good	38.9568	78.6484	4.6824	22.7124
Weak	48.0243	90.3021	7.2119	23.4616
[total]	86.9811	168.9506	11.8943	46.174
Ready4School				
NotReady	6.2312	19.2632	5.2429	5.2627
Ready	80.7499	149.6873	6.6514	40.9113
[total]	86.9811	168.9506	11.8943	46.1740

Time taken to build model (full training data) : 0.13 seconds

=== Model and evaluation on test split ===

EM

==

Number of clusters: 4

Number of iterations performed: 15

	Cluster			
	0	1	2	3
	(0.67)	(0.18)	(0.14)	(0.01)
Mother_educational_qualification				
Primary	1.0014	1.0048	1.0009	1.9929
Secondary	118.7411	2.5448	3.6229	1.0912
1st_degree	17.5305	32.1962	26.2287	1.0447
2nd_degree	1.0682	2.8270	1.1051	1.9997
MD	1.0190	2.1451	1.8354	1.0005
PhD	1.0000	1.0000	1.0000	1.0000
[total]	140.3602	41.7178	34.7929	8.1291
Father_educational_qualification				
Primary	1.0014	1.0048	1.0009	1.9929
Secondary	125.1342	4.8114	11.9522	1.1021
1st_degree	11.1159	32.1136	16.6646	1.1059
2nd_degree	1.0169	1.0269	1.0293	1.9269
MD	1.0918	1.7610	3.1459	1.0012
PhD	1.0000	1.0000	1.0000	1.0000
[total]	140.3602	41.7178	34.7929	8.1291
Mother_occupation				
UnEmployed	11.0942	3.4893	4.3554	2.0611
Private	113.1045	23.4636	15.4609	1.9710
Government	13.1615	11.7650	11.9766	1.0969
[total]	137.3602	38.7178	31.7929	5.1291
Father_occupation				
UnEmployed	1.9509	1.0013	1.0385	1.0094
Private	122.6440	30.0981	28.2693	2.9886
Government	12.7653	7.6184	2.4851	1.1311
[total]	137.3602	38.7178	31.7929	5.1291
Family_size				
Big	90.4762	33.3225	9.0878	3.1135
Small	45.884	4.3953	21.7051	1.0156
[total]	136.3602	37.7178	30.7929	4.1291
Child_position				
Late	22.4184	9.2881	1.2288	2.0647
Top	113.9418	28.4297	29.5641	2.0643
[total]	136.3602	37.7178	30.7929	4.1291
Good_Class_Peers				
Good	47.9633	9.1317	18.8523	3.0528
Weak	88.3969	28.5862	11.9406	1.0762
[total]	136.3602	37.7178	30.7929	4.1291
Weak_Class_Peers				
Good	70.3994	16.4449	7.0645	3.0912
Weak	65.9608	21.273	23.7284	1.0378
[total]	136.3602	37.7178	30.7929	4.1291
Good_ Neighbourhood _Peers				
Good	53.0704	15.6327	14.2677	2.0292
Weak	83.2898	22.0851	16.5252	2.0999
[total]	136.3602	37.7178	30.7929	4.1291
Weak_ Neighbourhood _Peers				
Good	67.2351	16.4299	19.2255	3.1095
Weak	69.1251	21.288	11.5674	1.0196
[total]	136.3602	37.7178	30.7929	4.1291
Ready4School				
NotReady	17.2763	2.5023	7.2169	2.0045
Ready	119.0839	35.2156	23.5760	2.1245
[total]	136.3602	37.7178	30.7929	4.1291

Time taken to build model (percentage split) : 0.05 seconds

Clustered Instances

0 75 (71%)
1 17 (16%)
2 11 (10%)
3 2 (2%)

Log likelihood: -6.85906

The ID3 algorithm WEKA results

=== Run information ===

Scheme: weka.classifiers.trees.Id3
Relation: Ready2Learn-weka.filters.unsupervised.attribute.Remove-R1-2,9-11,16-24
Instances: 306
Attributes: 11
Mother_educational_qualification
Father_educational_qualification
Mother_occupation
Father_occupation
Family_size
Child_position
Good_Class_Peers
Weak_Class_Peers
Good_Neighbourhood_Peers
Weak_Neighbourhood_Peers
Ready4School
Test mode: split 66.0% train, remainder test

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

ID3
Father_occupation = UnEmployed: NotReady
Father_occupation = Private
| Father_educational_qualification = Primary
| | Mother_occupation = UnEmployed: Ready
| | Mother_occupation = Private
| | | Mother_educational_qualification = Primary
| | | | Child_position = Late: NotReady
| | | | Child_position = Top: Ready
| | | Mother_educational_qualification = Secondary: Ready
| | | Mother_educational_qualification = 1st_degree: null
| | | Mother_educational_qualification = 2nd_degree: null
| | | Mother_educational_qualification = MD: null
| | | Mother_educational_qualification = PhD: null
| | Mother_occupation = Government: Ready
| Father_educational_qualification = Secondary
| | Mother_occupation = UnEmployed: Ready
| | Mother_occupation = Private
| | | Weak_Neighbourhood_Peers = Good
| | | | Mother_educational_qualification = Primary: null
| | | | Mother_educational_qualification = Secondary
| | | | | Good_Neighbourhood_Peers = Good
| | | | | | Family_size = Big: Ready
| | | | | | Family_size = Small
| | | | | | | Good_Class_Peers = Good
| | | | | | | | Weak_Class_Peers = Good: Ready

```
| | | | | | | | Weak_Class_Peers = Weak: Ready
| | | | | | | Good_Class_Peers = Weak: Ready
| | | | | Good_Neighbourhood_Peers = Weak
| | | | | | Good_Class_Peers = Good
| | | | | | | Family_size = Big
| | | | | | | | Child_position = Late: Ready
| | | | | | | | Child_position = Top
| | | | | | | | | Weak_Class_Peers = Good: Ready
| | | | | | | | | Weak_Class_Peers = Weak: Ready
| | | | | | | Family_size = Small: Ready
| | | | | | Good_Class_Peers = Weak
| | | | | | | Child_position = Late: Ready
| | | | | | | Child_position = Top
| | | | | | | | Family_size = Big
| | | | | | | | | Weak_Class_Peers = Good: Ready
| | | | | | | | | Weak_Class_Peers = Weak: Ready
| | | | | | | | Family_size = Small
| | | | | | | | | Weak_Class_Peers = Good: Ready
| | | | | | | | | Weak_Class_Peers = Weak: Ready
| | | | Mother_educational_qualification = 1st_degree
| | | | | Good_Neighbourhood_Peers = Good: NotReady
| | | | | Good_Neighbourhood_Peers = Weak: Ready
| | | | Mother_educational_qualification = 2nd_degree: null
| | | | Mother_educational_qualification = MD: null
| | | | Mother_educational_qualification = PhD: null
| | | Weak_Neighbourhood_Peers = Weak
| | | | Child_position = Late
| | | | | Weak_Class_Peers = Good
| | | | | | Mother_educational_qualification = Primary: null
| | | | | | Mother_educational_qualification = Secondary
| | | | | | | Good_Class_Peers = Good: Ready
| | | | | | | Good_Class_Peers = Weak
| | | | | | | | Good_Neighbourhood_Peers = Good: Ready
| | | | | | | | Good_Neighbourhood_Peers = Weak: Ready
| | | | | | Mother_educational_qualification = 1st_degree: Ready
| | | | | | Mother_educational_qualification = 2nd_degree: null
| | | | | | Mother_educational_qualification = MD: null
| | | | | | Mother_educational_qualification = PhD: null
| | | | | Weak_Class_Peers = Weak: Ready
| | | | Child_position = Top
| | | | | Weak_Class_Peers = Good
| | | | | | Family_size = Big
| | | | | | | Good_Class_Peers = Good
| | | | | | | | Good_Neighbourhood_Peers = Good: Ready
| | | | | | | | Good_Neighbourhood_Peers = Weak: NotReady
| | | | | | | Good_Class_Peers = Weak: Ready
| | | | | | Family_size = Small
| | | | | | | Good_Neighbourhood_Peers = Good
| | | | | | | | Good_Class_Peers = Good: Ready
| | | | | | | | Good_Class_Peers = Weak
| | | | | | | | | Mother_educational_qualification = Primary: null
| | | | | | | | | Mother_educational_qualification = Secondary: Ready
| | | | | | | | | Mother_educational_qualification = 1st_degree: Ready
| | | | | | | | | Mother_educational_qualification = 2nd_degree: null
| | | | | | | | | Mother_educational_qualification = MD: null
| | | | | | | | | Mother_educational_qualification = PhD: null
| | | | | | | Good_Neighbourhood_Peers = Weak: Ready
| | | | | Weak_Class_Peers = Weak
| | | | | | Good_Neighbourhood_Peers = Good
| | | | | | | Mother_educational_qualification = Primary: null
| | | | | | | Mother_educational_qualification = Secondary
| | | | | | | | Family_size = Big
| | | | | | | | | Good_Class_Peers = Good: Ready
| | | | | | | | | Good_Class_Peers = Weak: Ready
| | | | | | | | Family_size = Small: NotReady
| | | | | | | Mother_educational_qualification = 1st_degree
| | | | | | | | Family_size = Big: Ready
| | | | | | | | Family_size = Small: Ready
| | | | | | | Mother_educational_qualification = 2nd_degree: null
| | | | | | | Mother_educational_qualification = MD: null
| | | | | | | Mother_educational_qualification = PhD: null
| | | | | Good_Neighbourhood_Peers = Weak
| | | | | | Good_Class_Peers = Good: Ready
| | | | | | Good_Class_Peers = Weak
| | | | | | | Mother_educational_qualification = Primary: null
| | | | | | | Mother_educational_qualification = Secondary
| | | | | | | | Family_size = Big: Ready
| | | | | | | | Family_size = Small: Ready
| | | | | | | Mother_educational_qualification = 1st_degree: Ready
| | | | | | | Mother_educational_qualification = 2nd_degree: null
| | | | | | | Mother_educational_qualification = MD: null
| | | | | | | Mother_educational_qualification = PhD: null
| | Mother_occupation = Government
| | | Good_Neighbourhood_Peers = Good
| | | | Family_size = Big: Ready
| | | | Family_size = Small
| | | | | Weak_Class_Peers = Good: Ready
| | | | | Weak_Class_Peers = Weak
| | | | | | Weak_Neighbourhood_Peers = Good: NotReady
| | | | | | Weak_Neighbourhood_Peers = Weak
| | | | | | | Mother_educational_qualification = Primary: null
| | | | | | | Mother_educational_qualification = Secondary
| | | | | | | | Good_Class_Peers = Good: NotReady
| | | | | | | | Good_Class_Peers = Weak: Ready
| | | | | | | Mother_educational_qualification = 1st_degree
| | | | | | | | Good_Class_Peers = Good: Ready
| | | | | | | | Good_Class_Peers = Weak: NotReady
| | | | | | | Mother_educational_qualification = 2nd_degree: null
| | | | | | | Mother_educational_qualification = MD: null
| | | | | | | Mother_educational_qualification = PhD: null
| | | Good_Neighbourhood_Peers = Weak: Ready
| Father_educational_qualification = 1st_degree
| | Good_Class_Peers = Good
| | | Weak_Class_Peers = Good: Ready
| | | Weak_Class_Peers = Weak
| | | | Mother_occupation = UnEmployed
| | | | | Family_size = Big: NotReady
| | | | | Family_size = Small: NotReady
| | | | Mother_occupation = Private
| | | | | Weak_Neighbourhood_Peers = Good: Ready
| | | | | Weak_Neighbourhood_Peers = Weak
| | | | | | Mother_educational_qualification = Primary: null
| | | | | | Mother_educational_qualification = Secondary: NotReady
| | | | | | Mother_educational_qualification = 1st_degree
| | | | | | | Family_size = Big
| | | | | | | | Child_position = Late: Ready
| | | | | | | | Child_position = Top
| | | | | | | | | Good_Neighbourhood_Peers = Good: Ready
| | | | | | | | | Good_Neighbourhood_Peers = Weak: NotReady
| | | | | | | Family_size = Small: Ready
| | | | | | Mother_educational_qualification = 2nd_degree: null
| | | | | | Mother_educational_qualification = MD: null
| | | | | | Mother_educational_qualification = PhD: null
| | | | Mother_occupation = Government
| | | | | Family_size = Big: Ready
| | | | | Family_size = Small: NotReady
| | Good_Class_Peers = Weak
| | | Mother_educational_qualification = Primary: null
| | | Mother_educational_qualification = Secondary
| | | | Family_size = Big: Ready
```

```
| | | | Family_size = Small
| | | | | Weak_Class_Peers = Good: NotReady
| | | | | Weak_Class_Peers = Weak: Ready
| | | Mother_educational_qualification = 1st_degree: Ready
| | | Mother_educational_qualification = 2nd_degree: Ready
| | | Mother_educational_qualification = MD: Ready
| | | Mother_educational_qualification = PhD: null
| Father_educational_qualification = 2nd_degree
| | Mother_educational_qualification = Primary: null
| | Mother_educational_qualification = Secondary: null
| | Mother_educational_qualification = 1st_degree: null
| | Mother_educational_qualification = 2nd_degree: NotReady
| | Mother_educational_qualification = MD: null
| | Mother_educational_qualification = PhD: Ready
| Father_educational_qualification = MD: Ready
| Father_educational_qualification = PhD: null
Father_occupation = Government
| Weak_Neighbourhood_Peers = Good
| | Good_Class_Peers = Good: Ready
| | Good_Class_Peers = Weak
| | | Good_Neighbourhood_Peers = Good: Ready
| | | Good_Neighbourhood_Peers = Weak
| | | | Child_position = Late: NotReady
| | | | Child_position = Top
| | | | | Mother_educational_qualification = Primary: null
| | | | | Mother_educational_qualification = Secondary: Ready
| | | | | Mother_educational_qualification = 1st_degree
| | | | | | Mother_occupation = UnEmployed: NotReady
| | | | | | Mother_occupation = Private
| | | | | | | Father_educational_qualification = Primary: null
| | | | | | | Father_educational_qualification = Secondary: Ready
| | | | | | | Father_educational_qualification = 1st_degree: Not-
Ready
| | | | | | | Father_educational_qualification = 2nd_degree: null
| | | | | | | Father_educational_qualification = MD: null
| | | | | | | Father_educational_qualification = PhD: null
| | | | | | Mother_occupation = Government: null
| | | | | Mother_educational_qualification = 2nd_degree: null
| | | | | Mother_educational_qualification = MD: null
| | | | | Mother_educational_qualification = PhD: null
| Weak_Neighbourhood_Peers = Weak: Ready
```

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===
Time taken to test model on training split: 0 seconds

=== Summary ===

Correctly Classified Instances	86	82.6923 %
Incorrectly Classified Instances	17	16.3462 %
Kappa statistic	0.0278	
Mean absolute error	0.1612	
Root mean squared error	0.3523	
Relative absolute error	95.2445 %	
Root relative squared error	144.9646 %	
Coverage of cases (0.95 level)	89.4231 %	
Mean rel. region size (0.95 level)	57.6923 %	
UnClassified Instances	1	0.9615 %
Total Number of Instances	104	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.167	0.124	0.077	0.167	0.105	0.030	0.466	0.069	NotReady
	0.876	0.833	0.944	0.876	0.909	0.030	0.549	0.938	Ready
Weighted Avg.	0.835	0.792	0.894	0.835	0.862	0.030	0.544	0.888	

=== Confusion Matrix ===

a	b	<-- classified as
1	5	| a = NotReady
12	85	| b = Ready

The J48 algorithm WEKA results

=== Run information ===

Scheme: weka.classifiers.trees.J48 -U -M 4
Relation: Ready2Learn
Instances: 306
Attributes: 11
Mother_educational_qualification
Father_educational_qualification
Mother_occupation
Father_occupation
Family_size
Child_position
Good_Class_Peers
Weak_Class_Peers
Good_Neighbourhood_Peers
Weak_Neighbourhood_Peers
Ready4School
Test mode: split 66.0% train, remainder test

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

J48 unpruned tree

```
Weak_Neighbourhood_Peers <= 4
| Father_occupation = UnEmployed: No (1.0)
| Father_occupation = Private
| | Mother_educational_qualification = Primary: Yes (5.0/1.0)
| | Mother_educational_qualification = Secondary
| | | Weak_Neighbourhood_Peers <= 3: Yes (110.0/8.0)
| | | Weak_Neighbourhood_Peers > 3
| | | | Mother_occupation = UnEmployed: Yes (5.0)
| | | | Mother_occupation = Private
| | | | | Good_Neighbourhood_Peers <= 1: No (4.0/1.0)
| | | | | Good_Neighbourhood_Peers > 1
| | | | | | Family_size <= 4: Yes (9.0/1.0)
| | | | | | Family_size > 4
| | | | | | | Child_position <= 2
| | | | | | | | Good_Class_Peers <= 2: No (4.0/1.0)
| | | | | | | | Good_Class_Peers > 2: Yes (7.0/1.0)
| | | | | | | Child_position > 2: Yes (7.0/1.0)
| | | | Mother_occupation = Government: Yes (3.0/1.0)
| | Mother_educational_qualification = 1st_degree
| | | Weak_Class_Peers <= 4: Yes (60.0/3.0)
| | | Weak_Class_Peers > 4
| | | | Child_position <= 1: No (4.0/1.0)
| | | | Child_position > 1: Yes (12.0/2.0)
| | Mother_educational_qualification = 2nd_degree: No (1.0)
| | Mother_educational_qualification = MD: Yes (1.0)
| | Mother_educational_qualification = PhD: Yes (1.0)
| Father_occupation = Government: Yes (26.0/3.0)
Weak_Neighbourhood_Peers > 4: Yes (46.0)
```

Number of Leaves : 18
Size of the tree : 29
Time taken to build model: 0.01 seconds

=== Evaluation on test split ===
Time taken to test model on training split: 0 seconds

=== Summary ===

Correctly Classified Instances	95	91.3462 %
Incorrectly Classified Instances	9	8.6538 %
Kappa statistic	-0.0308	
Mean absolute error	0.1672	
Root mean squared error	0.2816	
Relative absolute error	94.1331 %	
Root relative squared error	109.2774 %	
Coverage of cases (0.95 level)	98.0769 %	
Mean rel. region size (0.95 level)	87.9808 %	
Total Number of Instances	104	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.000	0.021	0.000	0.000	0.000	-0.038	0.581	0.108	No
	0.979	1.000	0.931	0.979	0.955	-0.038	0.581	0.945	Yes
Weighted Avg.	0.913	0.934	0.869	0.913	0.891	-0.038	0.581	0.889	

=== Confusion Matrix ===

a	b	<-- classified as
0	7	a = No
2	95	b = Yes