Paper 1 "Benefits of Educational Data Mining "

Intro:

Today, most universities study the captain in data because anything that needs analysis wants to save its data and study on it, whether a practical or theoretical study. Data mining using data mining tools and methods helps in showing the correct result for the validity of the data. There are methods of prospecting in the educational environment, analyzing the methods under study, and coming up with percentages that are useful in making decisions in a particular environment, especially in an educational environment.

Educational Data Mining (EDM):

Previously, the data was downloaded and worked on manually, and that was useless or not extracting a feasible result from the data, and the data remains useless, but now, after working in data mining and data science, feasible results have been come out in various educational studies, whether coming out with psychological measurements or coming out by taking A decision to change a particular system.

The ultimate goal of EDM is to improve the educational process and To explain instructional strategies to make better decisions There are different definitions of EDM, but they all have in common that it is multidisciplinary An area of research that uses different methods and techniques from Machine learning, statistics, data mining, data analysis, to Analyze data collected during teaching and learning for I previously discovered information, relationships, and Patterns in large data warehouses.

The method is summed up in four stages, the first stage is translating the research problem into excavating the evidence, the second stage is the second stage, the search for data in the research problem, the third stage is the evaluation of the evidence, and the fourth stage is to come up with the results of mining the evidence.

Methods and algorithms are used in data mining, the most famous of which is classification, cluster analysis , prediction & association , regression analysis and decision trees .

Classification is a technique of data mining that is used to classify between classes and uses and trains the algorithm using pre-classified and that in tables based on some of characteristics as age, gender, grades, knowledge, academic achievements, motivation,

Cluster analysis is used to divide similar data into groups that were not previously specified. Useful in data processing Phase to define homogeneous groups used as Input for other models. Similar to classification and cluster analysis Can be used to investigate the similarities and differences between Students, courses and teachers .

Prediction refers to calculated assumptions for certain events made based on available processed data. Regression technique can be used for prediction, to model the relationship between one or more independent variables and dependent variables.

Association is a data mining technique used to discover the probability of the co-occurrence of items in a collection. The relationships between co-occurring items are called association rules.

Conclusion

Data mining in the educational process is a high-level work, and in our modern age it is considered one of the most influential sciences because it helps to extract secrets from data or hidden knowledge in the data.

Paper 2 "Data Mining of Family, School, and Society Environments Influences to Student Performance"

Intro:

It is common in the educational process that parents want their children to have their academic level and grades in continuous development and in this research the writer used data mining using methods in data mining using binary classification approach , the author has chosen decision tree and decision tree more accurate than neural network .

in research the author use data warehouse technique and get data from two sources , first from the journal and second from jakarta and tangerange city and total data 425 data , consists of 425 data and 395 from first source and 30 from data in second source in city .

Table 1. Attributes and Variables

A & V	Description
Sex	Student sex (binary) "F"(*female*), "M"(*male*).
Age	Student age (numeric) 15 until 22.
Address	Student's house location (binary) "U" (*urban*), "R"(*rural*).
Pstatus	Parents status (binary) T (*together*), A(*apart*).
Medu	Mother's education * (numeric) no education is 0, primary graduated is 1, Junior high graduated is 2, Senior high graduated is 3, Bachelor or diploma graduated is 4.
Mjob	Mother's job # (polynominal) education equals "*teacher*", medical equals "*health*", civil services (police, etc) equals "*services*", just at home equals "*at_home*", other equals "*other*".
Fedu	Father's education * (numeric no education is 0, primary graduated is 1, Junior high graduated is 2, Senior high graduated is 3, Bachelor or diploma graduated is 4.
Fjob	Father's job # (polynominal) education equals "*teacher*", medical equals "*health*", civil services (police, etc) equals "*services*", just at home equals "*at_home*", other equals "*other*".
Guardian	Main guardian of student (polynominal) "*father*", "*mother*", "*other*".
Famsize	Family size (binary) "GT3"(*greater than 3*), "LE3" (*less equal 3*).

Rapid Miner is a program licensed under the GNU Affero General Public License version

3[14]. This program has a convenient interface,

Figure 1. Decision Tree Result
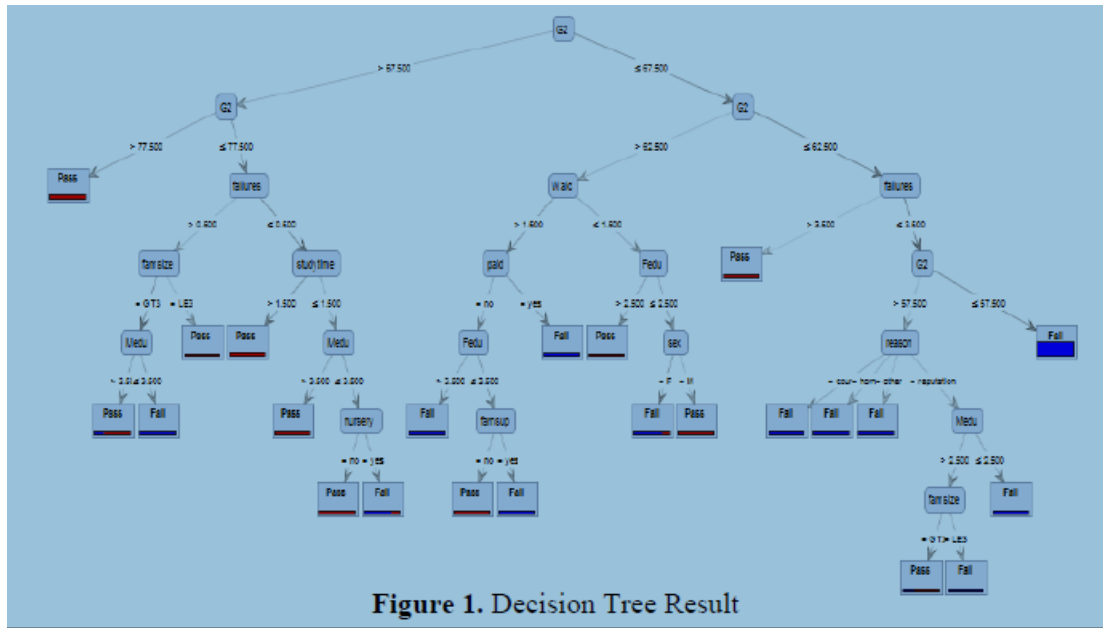
Result:

If the value of G2 is greater than 67.5, then in next branch it will be checked again, if G2 is greater than 77.5, then certainly the student will pass on the next test. If not, then the next check is whether failures (experiencing failure) is greater than 0.5 or not.

In other words, whether the student has experienced failure or never experienced it, because

the input received is only in integers 0.1,2,3,4. If the student has experienced a failure, then the next attribute, famsize (number of family members) will be examined. If famsize is a GT3 (Greater than 3 / more than 3), then Medu will be checked next (Mother's last education) if it is greater than 3.5 or in other words is 4 which means passing college, then the student is predicted to pass. If the mother's last education is below college graduation, then the student is predicted to fail or fail on the next test. Next, back to 2 node level above, if the famsize is LE3 (Less equal 3 / below or equal to 3 people) then the student is predicted to pass. Back to third level node from the top, if the student has never experienced a failure at all, then study time will be checked (out-of-school time). If the study time attribute of the student is more than 1.5 (at least studying with a minimum duration of 2 hours), then the student is predicted to graduate. However, if less than 1.5, then check the value of Medu, if Medu value is above 3.5 (at least must pass college), then the student is predicted to pass. If the Medu value is below or equal to 3.5, then the next attribute that will be checked is nursery (never entered nursery). If the value is a yes, then the student is predicted to fail, and if the value is a no, then the student is predicted to pass. Then return to the top level node, if the value of G2 is smaller than or equal to 67.5, then

G2 will be re-checked, whether G2 is greater than 62.5. If G2 is greater than 62.5, then the next attribute to be examined is Walc (the level of alcohol consumption at the weekend), whether it exceeds the level of 1.5 or not. If it exceeds the level 1.5, then the next attribute we check is paid (does the child attend a paid course outside school?).If the child is attending a paid course class, the child is predicted to fail, but if not then the

Fedu attribute (Father's last education) will be examined. If Fedu has value more than 3.5 (atleast graduated from college), then the child is predicted to fail, if the Fedu value is smaller or equal to 3.5, then check the famsup attribute (additional education by parent), if there is additional education from the parents, This child is predicted to fail, but otherwise, if this student does not get additional education from parents at home, then this child is predicted to pass.

Conclusion:

from 33 attributes and 425 rows of data which use them in study , it turns out that 12 of attributes really effect on result and achievements . upper secondary students. Those 12 attributes are sex, Medu, Fedu, famsize, reason, studytime, failures, famsup, paid, nursery, walc, and g2. 2nd year exam score (G2) is the most influential attribute compared to all other attributes. Both parents' education also has an important role to play in their children's achievement in school. Then failures (number of failures) also have an important effect. Learning from failures proves to play an important role in a child's achievement. Alcohol consumption also has a great effect on achievement. For other attributes, such as famsup, paid, nursery, sex, famsize, little effect on student achievement because of its position in the tree that are in the bottom of the root.

Paper 3 "Mining Big Data in Education: Affordances and Challenges"

Intro:

This paper I had taken which we will use it in our research which related with education and data mining.

We will talk about evaluating student knowledge using some application of data mining naming microlevel clickstream it based on sets of correct and incorrect responses to problem known as inference or latent knowledge estimation three methods are Bayesian Knowledge tracing , performance factors analysis and deep knowledge tracing , these methods help to know how this student level in knowledge .

Bayesian Knowledge tracing "BKT"

In this technique study when and how student be master in a unique skill we can use something like that in how can be a kid brilliant in some skills and make him specialized in it. and this technique mean its parameters like how make sure that this student master in some skill or old knowledge for that student related about this skill .

performance factors analysis "BFA"

parameters provide insight on the relative difficulty of skills and the relative learning associated with correct and incorrect answers. Extensions of PFA are an active area of research—for instance.

Deep Knowledge Tracing "DKT"

DKT has emerged as a popular alternative to BKT and PFA. DKT uses recurrent neural networks to model skill knowledge and mastery, producing a vector of the probability of mastery associated with each opportunity to practice a skill. Compared with the other approaches.

Conclusion:

The availability of big data offers exciting new threads of research and the opportunity to add additional perspective to existing threads in education. All types of big data in education offer affordances and challenges. The sheer amount of microlevel data make big data methods a powerful tool for analyzing learner processes, but that power can lead researchers to ignore broader and potentially more important patterns that cannot be measured at the microlevel. Macrolevel data provide a deep window into cognitive processes by examining individuals 'writing .

Paper 4 "A Predictive Model for Parental Stress and Constraining a Child's Own Activity: Structural and Data Mining Approach"

Intro:

The aim of the study was to test a theoretical model which described the causes of constraining a child's activity. The study involved 319 parents of preschool children. The model was tested with the help of structural equations. Cluster analysis was used to check how many clusters, i.e. groups of individuals similar to one another (due to variables described in the model), exist in the population. An artificial neural network was used to construct a prediction model for constraining children's activity. The results revealed that the theoretical model cannot be rejected as incorrect. The cluster analysis results revealed the existence of four groups of people. The neural network had a good prediction on constraining the activity of children.

Some authors predict that constraint of activity may also be associated with a decline in children's competence in the constrained areas [2] as research has confirmed. It turns out that constraining children's social activity is linked with the decline of social competence [3] and physical activity with increases of physical inactivity.

Types of constraining a child's activity:

When speaking of constraining the activity of a child we refer to the types of constraining activity and to the way in which it is constrained. The types of constraining children's activity comprise: a) constraint of manual activity, b) socio-relational, c) kinesthetic, d) musical, e) cognitive, etc. The methods of constraint include bans, criticism, ridiculing a child .

The problem of ethics in research on constraining children's activity:

Constraining a child's activity requires an analysis of the ethical aspects of this phenomenon. Some aspects of restraining have been banned legally; others, as research has revealed, have shown that they harm the child's development.

The alternative to constraining a child's activity:

Gurycka pointed to behaviors which may be an alternative to constraining children's activity. These behaviors are not associated with the increased likelihood of a mistake but, on the contrary, may have positive effects on the child's development. Gurycka termed these behaviors as directioning the children's activity. They are defined as creating conditions for the development of the child's own activity; agreeing on a substitutionary activity with the child negotiating. Directioning the child's activity is behavior on the part of the parent which creates situations that direct the child's activity and does not constrain it.

Cluster analysis:

The results of cluster analysis revealed that there are four clusters in the set . The first profile, which is the least frequently represented in the population, comprises 18% of the sample tested (Table 1), i.e. those who had the highest values in terms of discrepancies, experienced difficulties, children's representation, withdrawal, and constraining of children's activity. This group also has some of the highest scores in terms of pressure. Persons with

this profile used cognitive distancing the least often as a reaction to stress. The profile analysis provided new and valuable information. Approximately 54% of Polish parents does not experience the inability to meet educational goals; therefore, they do not experience difficulties and are not exposed to parental stress. About 27% of parents have a moderate level of difficulty in achieving their parental goals, and in this situation, they are characterized by a fairly diverse response to stress, including pressure, seeking help, and cognitive distancing. About 18% of parents, however, are exposed to higher stress in the parental situation due to the impossibility of achieving their parental goals. The profiling analysis also revealed that the greatest chance of obtaining positive reactions to stress, such as cognitive distancing or seeking help, occurs at low levels of stress. When stress increases, less positive forms of *stress coping, such as pressure and withdrawal, also occur. With high levels of difficulty, unfavorable forms of stress coping continue to strengthen while positive forms such as distancing and seeking help become weaker.*

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Discrepancy	128.291139	121.231884	244.164706	468.28
Difficulty	13.6329114	9.52173913	30.5411765	63.48
Representation	12.164557	9.33333333	21.6117647	33.48
Help seeking	22.9620253	8.50724638	15.0588235	15.76
Distancing	25.3164557	24.7681159	19.3058824	11
Pressure	5.70886076	4.79710145	15.1882353	15.96
Withdrawal	4.08860759	3	14.3294118	30.44
Constraining	29.164557	29.3478261	60,11,76,471	96.6
Number of cases	79	69	85	25
Percent (%)	30.620155	26.744186	32.9457364	9.68992248

Table 1: Number of cases classified in clusters.

Conclusion:

In conclusion, it should be noted that the formed latent variables showed very good reliability along with the percentage of variance extracted. Apart from the latent variable of the representation of the child, the reliability and percentage of variance extracted for the other variables had very good psychometric parameters. The model presented here has many degrees of freedom. This means that it is complex in the sense that it has many free parameters, not calculated. The value of the most important fit statistic RMSEA (<0.08) indicates that the model fits

the data well. The value of the test was, and thus the measurement model can be considered as fitting the data well. The value of CFI, which is slightly lower than 9, also shows that the model fits the data well.