

## **Big Data Analysis in School Adjustment Factors using Data Mining**

Sujeong Ko

*Dept. of Computer Software, Induk University, Korea*  
*sjko@induk.ac.kr*

### ***Abstract***

*Data mining technology is applied to various fields because it is a technique for analyzing vast amount of data and finding useful information. In this paper, we propose a big data analysis method that uses Apriori algorithm, which is a data mining technique, to find the related factors that have negative and positive influences on school adjustment. Among Korea Child and Youth Panel Survey(KCYPs), data related to adjustment to school life and data showing parental inclinations were extracted from the data of fourth grade elementary school students, first year middle school students, and high school freshman students, respectively and we have mapped the useful association rules among them. As a result, the factors affecting school adjustment were different according to the timing of the growth process, we were able to find interesting rules by looking for connections between rules. On the other hand, the factors that positively influenced school adjustment were not significantly different from each other, and overall, they were associated with positive variables.*

**Keywords:** *Apriori algorithm, Big data analysis, Data mining, School adjustment factors*

### **1. Introduction**

Among the super intelligent technologies in which the technology and the industry are intelligently advanced in the fourth industrial revolution era, a typical fusion technology is a technology related to the big data and the artificial intelligence [1]. Data mining is a big data analysis method combined with artificial intelligence and a text mining. The text mining techniques can extract useful information from informal text data and include emotional analysis, frequency analysis, and association analysis [2, 3].

Data mining technology is applied to various fields of data and extracts useful information to create new values [4]. In particular, it can create new useful values when applied to data with diverse characteristics, such as child or adolescent data. Finding the factors that can be associated with school adjustment or maladjustment among parental parenting schemes is a useful study to apply data mining techniques. Existing researches that studies the relationship between parents' parenting attitude and school adjustment are as follows. The first is a study on the influence of adolescent family structure on school adjustment through

multi-group analysis [5], and the second is a study on the relationship between parents, teachers, and peers in the relationship between self-efficacy and school adjustment [6], the last is a study on the parenting rearing attitude in the development trajectory relationship between adolescent school adjustment and academic achievement [7].

In this paper, we conducted association rule analysis using Apriori algorithm [8], which is a data mining technique, in order to find the factors that have negative and positive influences on the adjustment of school life in big data. Association rule analysis has the advantage of finding interesting links between very large numbers of variables and finding expert knowledge or interesting rules. For this purpose, the variables representing the students' school adjustment status and parental inclinations are extracted from KCYPS (Korea Child and Youth Panel Survey) [9], and we found out the specific and useful association rules from the data. In particular, we have analyzed longitudinal changes in the factors related to school adjustment as students grow up by the analysis of the factors affecting school adjustment for fourth grade in elementary school, first grade in middle school, and first grade in high school, respectively.

The composition of this paper is as follows. Section 2 describes Apriori algorithm for association rule mining. Section 3 analyzes the factors influencing school adjustment. Section 4 describes the performance evaluation, and Section 5 concludes.

## 2. Apriori algorithm for Association Rule Mining

### 2.1 Apriori Algorithm

To extract association rules from the database, Apriori algorithm has the advantage of not finding statistically calculated rules but finding unfamiliar rules. In other words, since association rule learning is self-learning, it does not train and does not need to be labeled. Since the infrequent item set in the data sets does not appear frequently in all subsets, it has the characteristic of a heuristic algorithm that reduces the number of rules by excluding it [10]. Figure 1 shows an R-program-based algorithm for finding association rules using the Apriori algorithm [11].

```
Association_rules <- apriori(data = dataset,
  parameter = list(support =  $\alpha$ , confidence =  $\beta$ ,
  min_length = 2)
```

- *dataset*: Subset(Korea Child and Youth Panel Survey)
- *apriori()*: Returns a rule object that stores all rules that meet the minimum criteria
- *Association\_rules*: The set association rules returned by the *apriori()* function
- *support*:  $\alpha$
- *confidence*:  $\beta$
- *min\_length*: The number of least associative variables

**Figure 1. Association rule mining using Apriori algorithm**

The algorithm generates rules from a set of items satisfying the minimum support threshold value ( $\alpha$ ) and the minimum confidence threshold value ( $\beta$ ). If the support threshold is set too high, the extracted rules are of common knowledge, and if the confidence threshold is set too high, rare item sets are pruned.

### 2.2 Support and Confidence

The support measures the frequency of how often an item set appears in a database [11]. Equation (1) defines the degree of support.

$$\text{Support}(X) = \text{count}(X) / N \quad (1)$$

In Equation (1),  $N$  is the number of students in the database, and  $count(X)$  is the frequency with which the variable  $X$  appears.

Confidence is a measure of predictive power and accuracy, and represents the ratio of the frequency of item sets, including both variables  $X$  and variables  $Y$ , among the set of items including variable  $X$ . Equation (2) defines the confidence.

$$Confidence(X \rightarrow Y) = Support(X, Y) / Support(X) \tag{2}$$

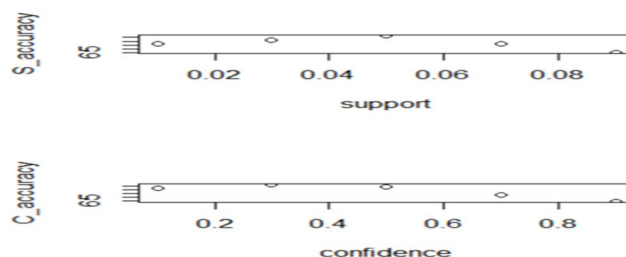
### 2.3 Threshold of Support and Confidence

In order to more precisely analyze the factors influencing school adjustment, five variables were selected from *KCYPs* data. Table 1 shows the five variables.

**Table 1. 5 variables related to school adjustment**

Variable name	Variable explanation	Variable segmentation	Variable value
EDU2A01	School adjustment (Learning activity)	School class time is interesting (Extract items with variable values 3 and 4)	1. Yes. 2. Somewhat. 3. It is not so. 4. Not at all.
EDU2A02		I do not miss school homework (Extract items with variable values 3 and 4)	
EDU2A05		I know what I learned in class (Extract items with variable values 3 and 4)	
EDU2A04		When there is something I do not know, I ask another person. (Extract items with variable values 3 and 4)	
EDU2A05		I do things in my study time (Extract items with variable values 1 and 2)	

The accuracy of the association rules varied according to the values of support and confidence. Figure 2 shows the accuracy of the results of mining association rules while changing the confidence and support based on the fourth grade 2,378 students of elementary school. In Figure 2, when the accuracy of mined association rules was the highest when the confidence was 0.3 and the support was 0.05.



**Figure 2. Accuracy by changing values of confidence and support**

## 3. Analyzing Influencing Factor of School Adjustment

### 3.1 Data Collection and Refinement

The data used to analyze the factors affecting school adjustment are the data from the fourth year of elementary school in 2010 to the first year of high school in 2016. Figure 3 shows the number of association rules that have adverse effects on school adjustment for each of the five variables in Table 1, designating a confidence level of 0.3 and a support level of 0.05 for fourth grade students of elementary school. In Figure 3, among the five variables related to school adjustment, the number of association rules extracted for the fourth variable in Table 1(EDU2A04) was very small. This is because this variable is not an interesting topic for students who cannot adjust to school. Therefore, the EDU2A04 variables in Table 1 are excluded from the analysis of adverse factors in school adjustment, and the other four variables are analyzed.

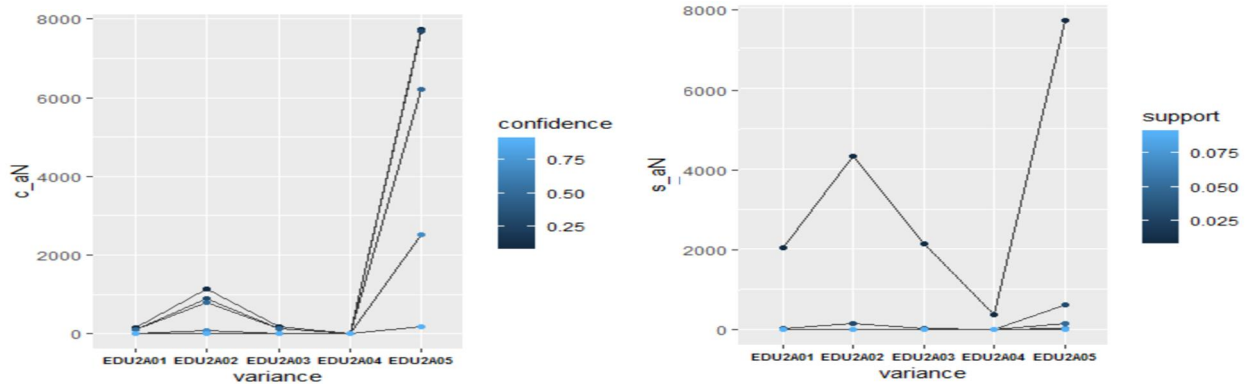


Figure 3. The number of association rules due to changes in confidence and support

In order to find out the variables related to school adjustment in Table 1, we also extract the values selected by the students for the 28 variables related to the students' psychological status and parenting attitude. Table 2 shows the descriptions and variable values of the 28 selected variables.

Table 2. 28 variables related to school adjustment

Variable name	Variable explanation	Variable segmentation	Variable value
HAK2A, HAK2B	Parental education	Father: middle or high school graduate 1,2 Mother: middle or high school graduate 1,2	1. Yes 2 That's it. 3 It is not so. 4 Not at all.
JAB1A, JOB1B	Parent job	Father has a Job: 1,2 Mom has a job: 1, 2	
PSY3B01, PSY3B02, PSY3B03	Life satisfaction	I am not happy, I have a lot of worries, my life is not happy.	
FAM2A01, FAM2A02, FAM2A03	Supervision	Parent knows where I go after school, how I spend my time, and when I get out.	
FAM2B01, FAM2B02, FAM2B03, FAM2B04	Affection	Parent respect for opinions, expression of liking, courage, praise.	
FAM2C01, FAM2C02, FAM2C03	Secret inertia	The inconsistency of the tongue, the feeling of being satisfied, the inconsistency depending on the presence of guests.	
FAM2D01, FAM2D02, FAM2D03, FAM2D04	Expect Excess	Expectations are burdensome, worrying, enthusiastic about studying, and saying better than others.	
FAM2E01, FAM2E02, FAM2E03, FAM2E04	Excessive interference	The degree of anxiety, emphasizing the need to win at all times, interfering with small things.	
FAM2F01, FAM2F02, FAM2F03	Rational explanation	Parent explain why I should do it, explain it before I go out, explain why it makes an unreasonable request.	

For the variables in Tables 1 and 2, the selected values evaluated by fourth grade elementary school students, first grade middle school students, and first grade high school students were followed up, and then as the grade increased, the data became more sparse. Figure 4 shows the density of the students evaluated for the variables in Tables 1 and 2. The (a) in Figure 4 shows the data of 4th grade in elementary school students, (b) shows the data of first grade in middle school students, and (c) shows the data of first grade in high school students. In Figure 4, as the grade rose it did not evaluate sincerely about the variables, therefore the density is 25.9% in (a), 21.35% in (b), and 18.23% in (c).

The variables not evaluated or evaluated as ‘NULL’ like JAB1A, JOB1B in Table 2 are irrelevant to the analysis for school adjustment. Therefore, they were excluded from the item set for association rule mining.

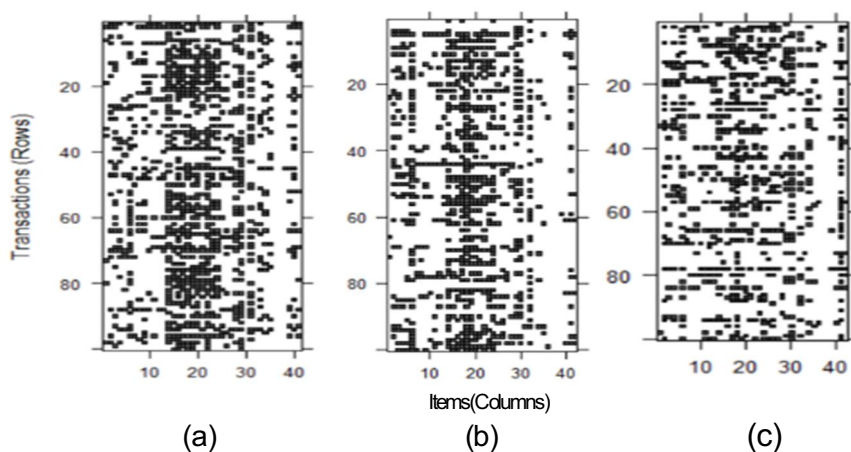


Figure 4. The density of the values for each variable evaluated by the students

### 3.2 Analysis of Related Factors Negatively Affecting School Adjustment

Applying preprocessed elementary fourth grade students’ data, first grade in middle school students’ data, first grade in high school students’ data to the Apriori algorithm in Figure 1, we can analyze the factors that have a negative effect on school adjustment. Table 3 shows a portion of the data set for mining the rules associated with EDU2A01 ("School Time is Fun"), a variable related to school adjustment. For the second student, the EDU2A01 responded to the question "Not so" or "Not", so it is included in the data set for mining association rules for EDU2A01 variable.

Table 3. Configuring transactions to mine association rules

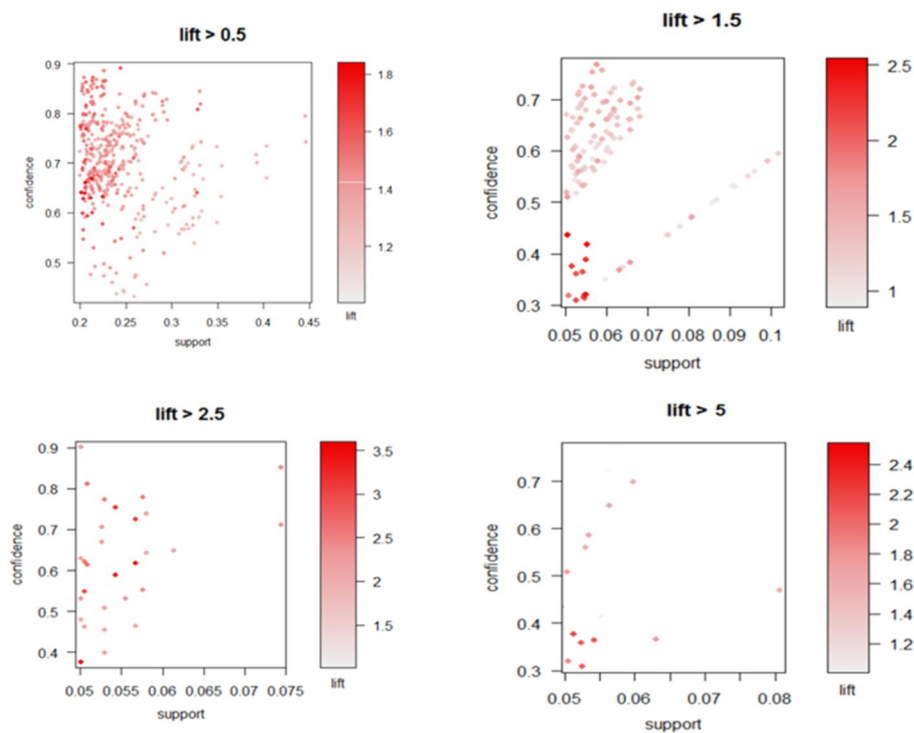
1	HAJ2A	HAJ2B	FAM2C03	FAM2C03	FAM2D01	FAM2D02	FAM2D03
2	HAJ2A	HAJ2B	PHY4A02	PSY3B02	FAM2C01	FAM2C02	<b>EDU2A01</b>
3	HAJ2A	HAJ2B	JOB1Bw1	PHY4A03	PHY4A04	PSY3B02	FAM2B02
4	HAJ2A	PHY4A02	PSY3B02	FAM2D01	FAM2D02	FAM2D03	FAM2D04
5	HAJ2A	HAJ2B	FAM2C03	FAM2F01	EDU2A01		

A data set composed of the form shown in Table 3 for each variable of EDU2A01, EDU2A02, EDU2A03, and EDU2A05 can be constructed, and association rules can be mined for the data of fourth grade elementary school students, first year students of middle school, and first grade students of high school. On the other hand, there is a disadvantage in that only the predictive association rules are extracted when association rules are mined using only the threshold of support and confidence. For example, assuming that an association rule in the field of purchasing goods in a store is to be mined, the association rule of "milk -> bread" will be mined. The conjecture rule that buy milk as such will also buy bread is not an interesting

analysis in the association factor analysis. In this case, the factor used is lift. Equation (3) is the equation for calculating the degree of lift [12].

$$Lift(X \rightarrow Y) = Confidence(X \rightarrow Y) / Support(Y) \quad (3)$$

The degree of lift of the Equation (3) is calculated by considering the relationship between the variable  $X$  and the variable  $Y$  when the variable  $Y$  is arbitrarily selected. The higher the value, the more interesting the rule is. Figure 5 shows that there is a difference in the number of extracted rules according to the change of the lift value. For the data of first grade of high school students, if we specify a support of 0.05, a confidence of 0.3, and a threshold of 0.5 for lift, then 114 rules, 1.5 for 52 rules, 2.5 for 26, and 5 Twelve rules were mined. Experimental results show that, when the threshold of the degree of lift is set to 1.5, the rule with the highest accuracy is mined.



**Figure 5. Generating association rules based on changes in lift value**

Table 4 shows an example of the association rule in fourth grade students of elementary school(E4), first grade students of middle school(M1), and first grade students of high school(E1). They are mined from related variables that negatively affect school adjustment for four variables. In Table 4, the mining results for EDU2A01 as  $\{FAM2E04w1, PSY3B02w1\} = \{EDU2A01w1\}$  indicate that students who do not have fun at school are more likely to "Not let parent do what I want to do" indicating that there is a strong correlation between the two variables.

We can analyze the causes of school maladjustment based on the association rules extracted in the form of Table 4. As a result of analyzing the association rules as shown in Table 4, as the number of students grows and the environment changes, the cause of school maladjustment also changes in 4th grade students of elementary school, first grade students of middle school, and first three years students of high school. Figure 6 shows the relationship between the variables of EDU2A01 and 26 variables. As shown in (a), (b), and (c)

in Figure 6, each variable is affected by several factors rather than being influenced by specific variables. Also, it shows that the cause of school maladjustment is changing significantly as the grade increases.

**Table 4. Examples of mining association rules associated with maladjustment**

	Variable	Association rule	Support	Confidence	Lift
E4	EDU2A01	{FAM2E04w1,PSY3B02w1} => {EDU2A01w1}	0.05046	0.31830	1.8552
	EDU2A02	{EDU2A02w1,FAM2C03w1,FAM2D04w1,FAM2E04w1} => {FAM2E03w1}	0.05088	0.76100	2.0945
	EDU2A03	{EDU2A03w1,FAM2D04w1} => {FAM2E03w1}	0.05046	0.59113	1.6269
	EDU2A05	{EDU2A05w1,FAM2C01w1,FAM2C02w1,FAM2C03w1,FAM2D04w1,FAM2E02w1,FAM2E04w1} => {FAM2E03w1}	0.05298	0.90000	2.4770
M1	EDU2A01	{EDU2A01w1,FAM2F02w1} => {FAM2F03w1}	0.05172	0.53947	3.1754
	EDU2A02	{EDU2A02w1,FAM2D04w1,FAM2E04w1} => {FAM2E03w1}	0.05004	0.71686	2.1939
	EDU2A03	{EDU2A03w1,FAM2E03w1,FAM2E04w1} => {FAM2C02w1}	0.05088	0.69142	2.2189
	EDU2A05	{EDU2A05w1,FAM2B04w1} => {FAM2B02w1}	0.05088	0.48594	3.7156
H1	EDU2A01	{EDU2A01w1,FAM2F02w1} => {FAM2F01w1}	0.05677	0.61643	3.4090
	EDU2A02	{EDU2A02w1,HAJ2Aw1} => {HAJ2Bw1}	0.05719	0.80473	2.2620
	EDU2A03	{EDU2A03w1,FAM2F03w1} => {FAM2F02w1}	0.05088	0.71597	3.0134
	EDU2A05	{EDU2A05w1,FAM2E03w1,FAM2E04w1} => {FAM2C03w1}	0.05004	0.61658	3.4827

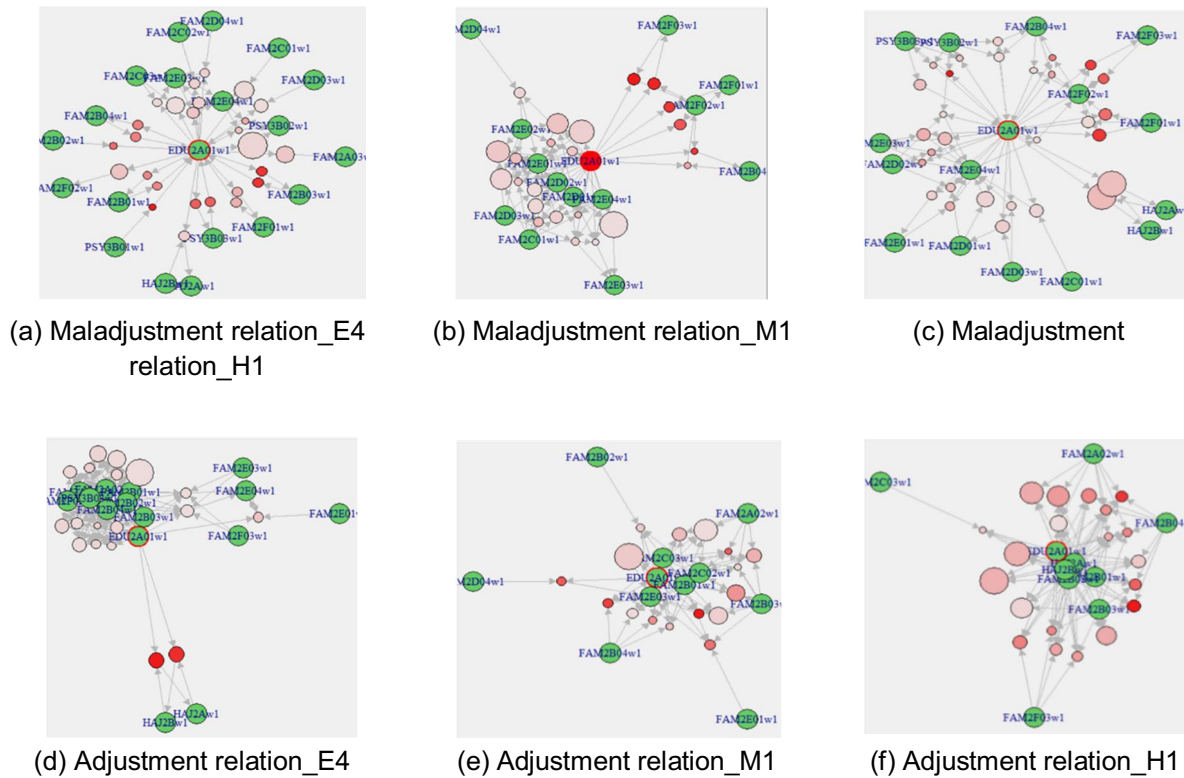
**3.3 Analysis of Affecting Factors Affecting School Adjustment Positively**

In order to analyze the factors affecting school adjustment positively, we construct a data set based on the variables that positively answered the variables in Table 2. The results are highly dense compared to the data sets with the negative impacts of Table 3. In addition, the data set negatively affecting adjustment excluded the variable of EDU2A04 ("Ask others when there is something I do not know"), while students in the set related to positive adjustment is do not care about the variable of EDU2A05("I do things in my study time"), so we exclude the variable from the analysis. If the association rules are mapped after specifying the confidence, support, and degree of lift for data sets thus constructed, the association rules are mapped in Table 5.

**Table 5. Examples of mining association rules associated with school adjustment**

	Variable	Association rules	Support	Confidence	Lift
E4	EDU2A01	{EDU2A01w1,FAM2A03w1,FAM2B02w1,FAM2F03w1}=>{FAM2F02w1}	0.50630	0.91489	1.1709
	EDU2A02	{EDU2A02w1,FAM2A02w1,FAM2B01w1,FAM2B02w1,FAM2B03w1,FAM2E04w1} => {FAM2E03w1}	0.30445	0.86293	1.3571
	EDU2A03	{EDU2A03w1,FAM2D04w1} => {FAM2E03w1}	0.05046	0.59113	1.6269
	EDU2A04	{EDU2A04w1,FAM2B03w1,HAJ2Bw1} => {HAJ2Aw1}	0.30151	0.85971	1.6122
M1	EDU2A01	{EDU2A01w1,FAM2F02w1} => {FAM2F03w1}	0.05172	0.53947	3.1754
	EDU2A02	{EDU2A02w1,FAM2B02w1} => {FAM2A03w1}	0.50105	0.89063	1.2395
	EDU2A03	{EDU2A03w1,FAM2B01w1,FAM2B02w1} => {FAM2B04w1}	0.50420	0.92159	1.2944
	EDU2A04	{EDU2A04w1,FAM2F02w1} => {FAM2F03w1}	0.50903	0.91673	1.2912
H1	EDU2A01	{EDU2A01w1,FAM2A02w1,FAM2B01w1} => {FAM2B03w1}	0.50294	0.92857	1.2867

EDU2A02	{EDU2A02w1,FAM2B04w1} => {FAM2B03w1}	0.50294	0.9568	1.3259
EDU2A03	{EDU2A03w1,FAM2B01w1} => {FAM2B02w1}	0.50756	0.89275	1.2850
EDU2A04	{EDU2A04w1,FAM2B02w1} => {FAM2B04w1}	0.51303	0.88214	1.3665



**Figure 6. The relationship between variables associated with the variable of EDU2A01**

In Table 5, the rules that adversely affect school adjustment differ according to the variables of EDU2A01, EDU2A02, EDU2A03, and EDU2A05. As shown in (d), (e), and (f) in Figure 6, similar results were also obtained when the analysis was done according to the period of the fourth grade of elementary school(E4), the first grade of middle school(M1), and the first grade of high school(H1). In other words, it is associated with common variables such as "Parent like us", "Parent give me courage when I get tough", "Parent explain why they cannot do it when they make an unreasonable request" and "Parent respect my opinion".

**4. Performance Evaluation**

In order to evaluate the performance of the method of analyzing the school adjustment influencing factors using data mining, we compared this method using frequency analysis (F\_analysis) [13] and cluster analysis method (C\_analysis) [14] used in a text mining of big data, respectively. For frequency analysis, we use inverse document frequency (TF · IDF-Term Frequency) and K-means clustering method [15]. K-means clusters are formed by setting the number of desired clusters around the initial value, and classifying each data into closest clusters.



#### 4.1 Evaluation Data

Randomly we selected 1000 students from fourth grade elementary school data in the KCYPS. Among the 1,000 students, 200 students were excluded from the evaluation data for the variables shown in Table 1 and Table 2, with 'NULL' as the basis for the answers given in first grade middle school and first grade high school as well as in fourth grade elementary school.

#### 4.2 Evaluation Measure

F-measure [15] is used as an evaluation measure to evaluate the performance of the big data analysis method in school adjustment factors. The F-measure measures the performance of the analysis results using precision and recall, and Equation (4) represents the precision and Equation (5) represents the recall.

$$P = \frac{N_{Real} \cap N_{Association}}{N_{Association}} \quad (4)$$

$$P = \frac{N_{Real} \cap N_{Association}}{N_{Real}} \quad (5)$$

In Equation (4) and Equation (5),  $N_{Real}$  represents the number of students who did not actually achieve adjustment to the variables shown in Table 1, and  $N_{Association}$  represents the number of students classified as not adjustable based on association rules. Applying precision and recall can be expressed as Equation (6).

$$F1 = \frac{2PR}{P+R} \quad (6)$$

In Equation (6), F1 score reaches its best value at 1 and worst at 0 and we apply the same weighting to accuracy and recall.

#### 4.3 Performance Evaluation Result

Figure 7 shows the precision and recall of the TF-IDF, K-means, and Apriori methods. In Figure 7, the Apriori method shows high performance in both precision and recall, but the K-means method has high recall but low precision. The precision of the TF-IDF method is higher than the K-means method, but the recall is lower than the method.

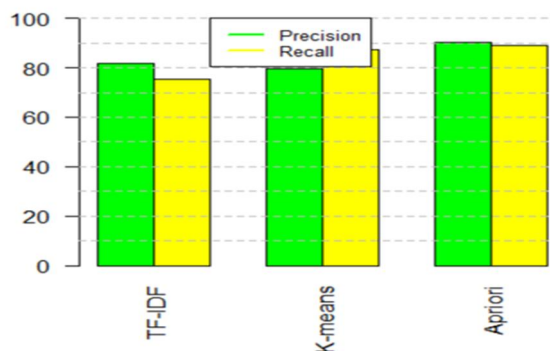


Figure 7. Precision and recall for methods of TF-IDF, K-means, Apriori

Figure 8 shows the F1 values of the TF-IDF, K-means, and Apriori methods while increasing the number of students from 100 to 800. The method using K-means has lower performance than other methods when the number of users belonging to the cluster is small. The TF-IDF method is less sensitive to the number of students than the K-means method, but the overall performance is low because the variables are analyzed using frequency only. On the other hand, the Apriori method has higher performance than other methods because it uses mining rules of related variables.

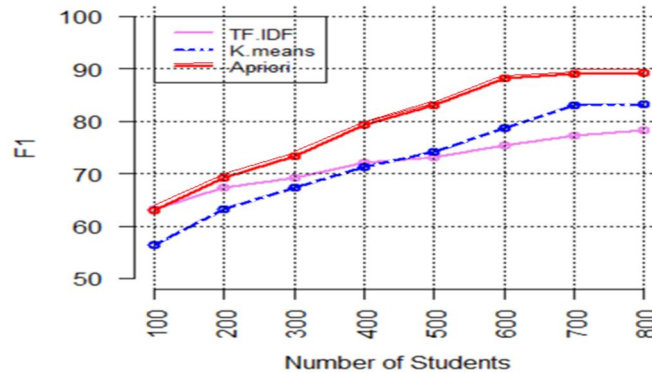


Figure 8. F1 Value for methods of TF-IDF, K-means, Apriori

## 5. Conclusion

In this paper, we conducted a big data analysis using data mining to find out the factors that have negative influences and the factors that have positive influences on school adjustment. Four variables related to school adjustment were selected and the association rules were mined to find out which variables were correlated with to school life and parental inclinations. As a result, the factors positively affecting the school adjustment were different according to the time of the growth process for each of the four variables. On the other hand, the factors negatively influencing school adjustment were not significantly different in different growth periods, and that they are generally associated with positive variables such as expressing "Parent like me", "Parent give me courage when I get hard," and explaining "Why not when parent make unreasonable demands."

In the future, it is necessary to investigate factors affecting school adjustment using data mining methods other than Apriori algorithm.

## References

- [1] J. R. Finkel, T. Grenager, and C. Manning. "Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling," in *Proc. of the 43rd Annual Meeting of the Association for Computational Linguistics*, 2005. URL:<http://aclweb.org/anthology/P05-1045>.
- [2] Sandoval, A. M., and T. Redondo, "Text Analytics: the convergence of Big Data and Artificial Intelligence," *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 3, No. 6, 2016. DOI: <http://dx.doi.org/10.9781/ijimai.2016.369>.
- [3] D. Boyd, and K. Crawford, "Six Provocations for Big Data," *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, 2011. DOI: <http://dx.doi.org/10.2139/ssrn.1926431>.
- [4] J. Chang, "An Experimental Evaluation of Box office Revenue Prediction through Social Big data Analysis and Machine Learning," *The Journal of The Institute of Internet, Broadcasting and Communication*, Vol. 17, Issue 3,

- pp. 167-173, 2017.  
DOI: <https://doi.org/10.7236/JIIBC.2017.17.3.167>.
- [5] J. KIM, Y. Jang, J. Min, "A Study on the Effect of School violence to Adolescent's School Adjustment : Moderating Effect of Parent-child Communication," *Korean Journal of Youth Studies*, Vol. 18, No. 7, pp. 209-234, 2011. UCI:G704-000387.2011.18.7.005.
- [6] B. Khu, "The Mediation Effects of relationship with parent, teacher, and peer between Self-efficacy and Adjustment to School," *Korean Journal of Youth Studies*, Vol. 19, No. 3, pp. 347-373, 2012. UCI: G704-000387.2012.19.3.010.
- [7] J. Kim, "The Longitudinal Relationship between School adjustment and Academic achievement in Adolescents on the Parenting attitude," *The Journal of Counseling. Korean Counseling Association (KCA)*, Vol. 17, No. 2, pp. 303-326, 2016.  
DOI: 10.15703/kjc.17.2.201604.303.
- [8] Agrawal, R. and Srikant, R. "Fast Algorithms for Mining Association Rules in Large Databases," in *Proc. of the 20th International Conference on Very Large Data Bases*, pp. 487-499, 1994. ISBN:1-55860-153-8.
- [9] National Youth Policy Institute, *1st 7th Survey Data User's guide in Korea Children and Youth Panel Survey(KCYPS)*, National Youth Policy Institute, Seoul, 2017. URL: <http://archive.nypi.re.kr>.
- [10] Y. Kim, W. Kim, and U. Kim, "An Efficient Method for Mining Frequent Patterns based on Weighted Support over Data Streams," *The Journal of Korea Academia-Industrial cooperation Society*, Vol. 10, No. 8, pp. 1998-2004, 2009. UCI: G704-001653.2009.10.8.049.
- [11] Brett Lantz, *Machine Learning with R Kindle Edition*, Packt, pp. 324-326, 2013. ASIN: B00G9581JM.
- [12] Nada Hussein, Abdallah Alashqur, and Bilal Sowan, "Using the interestingness measure lift to generate association rules," *Journal of Advanced Computer Science & Technology*, Vol. 4, No. 1. pp. 156-161, 2015.  
DOI: <http://dx.doi.org/10.14419/jacst.v4i1.4398>.
- [13] Qaiser, Shahzad and Ali, Ramsha, "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents," *International Journal of Computer Applications*, Vol. 181, No. 1, 2018.  
DOI: <https://10.5120/ijca2018917395>.
- [14] S. Oh, "Design and Analysis of TSK Fuzzy Inference System using Clustering Method," *The Journal of Korea Institute of Information, Electronics, and Communication Technology*, Vol. 7, No. 3, pp. 132-136, 2014. UCI: G704-SER000003092.2014.7.3.004.
- [15] Aparna Upadhyay, Ravindra Gupta, and Varsha Namdev, "Clustering analysis based learning of Web Mining," *International Journal of Advance Engineering and Research Development*, Vol. 4, Issue 6, 2017. e-ISSN: 2348 - 4470, print-ISSN: 2348-6406.