

# Data Mining Classification of Intelligence Quotient in High School Students

*by* Fakultas Teknik

---

**Submission date:** 19-Jan-2024 02:37PM (UTC+0700)

**Submission ID:** 2273774624

**File name:** Prosiding\_5.pdf (828.46K)

**Word count:** 2718

**Character count:** 14360

# Data Mining Classification of Intelligence Quotient in High School Students

1<sup>st</sup> Des Suryani  
Department of Informatics Engineering  
Universitas Islam Riau  
Pekanbaru, Indonesia  
[des.suryani@eng.uir.ac.id](mailto:des.suryani@eng.uir.ac.id)

2<sup>nd</sup> Ause Labellapansa  
Department of Informatics Engineering  
Universitas Islam Riau  
Pekanbaru, Indonesia  
[ause.labella@eng.uir.ac.id](mailto:ause.labella@eng.uir.ac.id)

3<sup>rd</sup> Setia Wulandari  
Department of Informatics Engineering  
Universitas Islam Riau  
Pekanbaru, Indonesia  
[setiawulandari@student.uir.ac.id](mailto:setiawulandari@student.uir.ac.id)

4<sup>th</sup> Ahmad Hidayat  
Department of Psychology  
Universitas Islam Riau  
Pekanbaru, Indonesia  
[ahmadhidayat@psy.uir.ac.id](mailto:ahmadhidayat@psy.uir.ac.id)

**Abstract**—An educational institution of “SMA Negeri 1 Padang” has been conducting psychological tests using the Intelligenz Structure Test (IST) in determining the classification of Intelligence Quotient (IQ) of their own students high school. Up to this day classification process of student IQs has been handled manually by psychologists and took an average of 15 days. This study applies data mining study which is a series of processes of getting knowledge or pattern from data set using method of classification that is C4.5 algorithm, which is useful to simplify and save time for psychologists. This study consisted of 10 criteria involving 270 data and resulted in output in the form of four kinds of student IQ classification consisting of very superior, superior, above average and average categories. Based on the test results classification of student IQs is obtained as 84%.

**Keywords**- Intelligenz Structure Test, Intelligence Quotient, C4.5 Algorithm

## I. INTRODUCTION

An educational institution of SMA Negeri 1 (State Senior High School) Padang has been conducting psychological tests using IST as a psychology test in determining the classification of their student IQs. The IST test used is a Rudolf Amthauer instrument which consists of 10 subtests: age, completing sentence (SE), word equation (WA), verbal analogy (AN), character similarity (GE), numerical ability (RA), number series (ZR), selecting image (FA), cube drill (WU), and memories (ME) and to determine the scale of the IQ student classification, they use David Wechsler's scale which consists of 7 categories: highly superior, superior, above average, average, below average, borderline and mentally retarded. Up to this day classification process of students IQs has been handled manually by psychologists and took an average of 15 days. This study applies data mining study which is a series of processes of getting knowledge or pattern from data set using method of

classification algorithm C4.5, which is useful to simplify and save time for psychologists.

The growth of information technology is so much advanced today, causing the level of accuracy of a data is highly needed in everyday life. Each information that disseminates becomes an important thing to determine on every decision in a particular situation. This leads to the provision of information to be analyzed and summarized into a knowledge of useful data when a decision is made, ie data mining.

Data mining is a set of processes to gain knowledge or patterns from data set using a particular technique or method. The classification method used is the Decision Tree, which is one of the easy classification method for human interpretation. The structure resembles a tree, where the node signifies an attribute, the branch is the value of the attribute and the leaf is the class. The algorithm used is C4.5 algorithm. C4.5 algorithm is used to form decision trees. This algorithm can also simplify the knowledge of the system so that the inference process can be faster [1].

The results of the study and modeling process using Decision Tree C4.5 algorithm conducted on the data of contraceptive use in Pratiema Hasanah Pekanbaru clinic can be drawn conclusion that the results of training data study of contraceptive usage tools can form knowledge in the form of decision tree models that transform data into decision trees that represent the rules. Accuracy rate obtained from data processing reach 93.15% which can be classified Excellent so that this result can be a reference for midwife in advise patient in choosing contraception [2].

## II. LITERATURE REVIEW

### A. Intelligence Quotient

Intelligence quotient (IQ) is a collection or the totality of a person's ability to act in a certain purpose, think rationally, and face the surrounding environment effectively. One of the most common ways used to express high IQ level is to translate the

intelligence test results into numbers that can be indicative of a person's IQ level when relatively compared to a norm [3].

### B. Intelligence Structure Test (IST)

IST test is one of the psychological tests used to measure individual intelligence [4]. IST consists of nine subtests totaling 176 items. The nine subtests in the IST are Satzerganzung (SE) completing the sentence, Wortauswahl (WA) word equation, Analogien (AN) verbal analogy, Gemeinsamkeiten (GE) character similarity, Rechenaufgaben (RA) numerical ability, Zahlenreihen (ZR) number series, Figurenauswahl (FA) selecting image, Wurfelaufgaben (WU) cube drills, and Merkaufgaben (ME) memories [5].

TABLE 1. CLASSIFICATION OF IQ BASED ON DAVID WECHSLER SCALE [5]

IST	IQ	Category
≥ 120	≥ 130	Very superior
113 – 119	120 – 129	Superior
107 – 112	110 – 119	Above average
93 – 106	90 – 109	Average
87 – 92	80 – 89	Below average
80 – 86	70 – 79	Borderline
≤ 79	≤ 69	Mentally retarded

### C. Data Mining

Data mining is a series of processes to gain knowledge or patterns from data set [6]. Data Mining solves the problem by analyzing the data which is already in the database. Data Mining, often called knowledge discovery in databases (KDD), is an activity that includes data collection, usage of historical data to find patterns of regularity and relationship patterns in large data sets [7]. The output of data mining can be used to improve future decision making.

### D. Classification

The classification technique is a systematic approach to construct a classification model of an input data set [8]. Classification is the learning process of a goal or a target function that maps each set of  $x$  attributes to one of the previously defined  $y$  labels. The target function is also called the classification model.

### E. Decision Tree

The decision tree is used as a reasoning procedure to get answers to the problems that are entered [6]. The tree formed is not always a binary tree. If all the features in the data set use 2 kinds of categorical values then the tree form shown is a binary tree. If the feature contains more than 2 kinds of categorical values or uses numeric type then the form of the tree shown is usually not a binary tree.

### F. C4.5 Algorithm

The C4.5 algorithm was introduced by Quinlan in 1995 as an improved version of ID3. Generally the C4.5 algorithm for decision tree building is as follows [6]:

1. Select an attribute as root.
2. Create a branch for each value.
3. Distributing cases in the branch.
4. Repeat the process for each branch until all the cases on the branch have the same class.

Entropy is used to determine how informative an entered attribute is to generate an attribute. The basic formula of entropy is as follows [6]:

$$\text{Entropy}(S) = \sum_{i=1}^k -p_i * \log_2 p_i \quad (1)$$

Information :

$S$  = case set.

$k$  = number of  $S$  partitions.

$p_i$  = The proportion of  $S_i$  to  $S$ .

To select an attribute as a root, based on the highest gain value of the attributes. To calculate the gain, the following formula given is as follows [6]:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (2)$$

Information :

$S$  = Case set.

$A$  = Attribute.

$k$  = Number of attribute partitions  $A$ .

$|S_i|$  = Number of cases on  $i$  partition.

$|S|$  = Number of cases in  $S$

### G. Accuracy

A classification system is expected to classify all data sets correctly. However, it can not be denied that the performance of a system can not work 100% correctly [6]. Therefore, a classification system should also be measured for performance. Generally measuring classification performance uses confusion matrix. The confusion matrix is a table that records the results of classification work.

$$\text{Accuracy} = \frac{\text{the amount of data predicted correctly}}{\text{number if predictions performed}} \quad (3)$$

## III. RESEARCH METHODOLOGY

### A. Data Collecting

The data collection techniques required in the classification of IQs, obtained from interviews and literature study.

- Interviews, conducted to collect information that would be useful in classifying student IQs. The interview was conducted on psychologists who organize psychology tests in SMA Negeri 1 Padang.
- Literature study, collecting data by searching and studying from various sources related to the problems studied in the preparation of this thesis, both from

books, scientific journals and from other readings that can be accounted for.

#### IV. RESULTS AND DISCUSSIONS

From the explanation of the analysis and the design that has been done in the previous chapter related, the next stage is to implement and test the system by testing black box which is a test that focuses on the functional requirements of the software. Thus, enabling software engineers to obtain a set of input conditions that fully utilize all functional requirements of a program. Then by testing end users to get good results leading to the final objective of the initial plan of the system.

##### A. C4.5 Algorithm Testing

On the C4.5 algorithm testing using 270 student data, the training data is used to classify IQ classes consisting of 4 categories namely very superior ( $\geq 130$ ), superior (120 - 129), above average (110 - 119), and average (90 - 109). There are 10 criteria used for IQ classification with C4.5 algorithm, with details as follow:

- WA criteria (word equation), consisting of 6 categories that are very superior ( $\geq 120$ ), superior (113 - 119), above average (107 - 112), average (93 - 106), below average (87 - 92), and borderline (80 - 86).
- AN Criteria (verbal analogy), consisting of 6 categories such very superior ( $\geq 120$ ), superior (113 - 119), above average (107 - 112), average (93 - 106), below average (87 - 92), and borderline (80 - 86).
- ME criteria (memory), consisting of 6 categories which are very superior ( $\geq 120$ ), superior (113 - 119), above average (107 - 112), average (93 - 106), below average (87 - 92), and borderline (80 - 86).
- GE criteria (the same nature), consisting of 6 categories with very superior ( $\geq 120$ ), superior (113 - 119), above average (107 - 112), average (93 - 106) (87 - 92), and borderline (80 - 86).
- RA criterion, consisting of 6 categories, they are very superior ( $\geq 120$ ), superior (113 - 119), above average (107 - 112), average (93 - 106), below average (87 - 92), and borderline (80 - 86).
- WU criteria (cube exercises), consisting of 6 categories: very superior ( $\geq 120$ ), superior (113 - 119), above average (107 - 112), average (93 - 106), below average (87 - 92), and borderline (80 - 86).
- ZR criteria (series of numbers), consisting of 6 categories: very superior ( $\geq 120$ ), superior (113 - 119), above average (107 - 112), average (93 - 106), below average (87 - 92), and borderline (80 - 86).
- FA Criteria (selecting images), consisting of 6 categories: very superior ( $\geq 120$ ), superior (113 - 119), above average (107 - 112), average (93 - 106), below average (87 - 92), and borderline (80 - 86).
- SE criteria (complete sentence), consists of 6 categories: very superior ( $\geq 120$ ), superior (113 - 119), above average (107 - 112), average (93 - 106), below average (87 - 92), and borderline (80 - 86).

- Age criteria, consisting of 4 categories ie age 13, age 14, age 15, and age 16.

##### 1. Entrophy Value Calculation

The foremost step of the C4.5 algorithm is to find the entropy value. First is to determine the total entropy value in the case. The entropy counts on each criterion. Next is to calculate the gain value for each criterion.

##### 2. Gain Value Calculation

After all entropy calculations on each criterion are completed, then the gain value is calculated. The highest gain value of 0.3503 is made as the first root node. For the next root node, calculation will be done again until all formed patterns are met.

##### 3. Determining the Root

From the calculation to find the value of entropy and gain value, it can be identified that attribute with highest gain is ZR which is equal to 0.3575. Therefore ZR becomes the root node. The data will then be broken down into 6 groups: "very superior", "superior", "above average", "average", "below average", and "borderline". The following decision tree is formed for node 1:

##### B. System Testing

##### 1. Value input of each criterion

Field	Value
Registration Number *	9 / IQ / VIII / 2018
NISN *	3456789012
Student Name *	Dewi Sartika
Test Date *	24-08-2018
Word Equation (WA) *	124
Verbal Analogy (AN) *	104
Memories (ME) *	124
Character Similarity (GE) *	115
Numerical Ability (RA) *	84
Cube Drill (WU) *	116
Number Series (ZR) *	114
Selecting Image (FA) *	115
Completing Sentence (SE) *	94
Age *	14

Fig 1. Criterion Value Input

From Fig 1, it can be explained that the input of criterion value is the same as the criterion in manual test.

## 2. High School Students 'IQ Classification Output

Field	Value
Registration Number	919219012018
NISN	3456789012
Test Date	24.08.2018
Word Equation (WA)	124 (Highly Superior)
Verbal Analogy (VA)	124 (Average)
Memories (ME)	124 (Highly Superior)
Character Similarity (CS)	115 (Superior)
Numerical Ability (NA)	94 (Disturbance)
Cube Drill (CD)	119 (Superior)
Number Series (NS)	114 (Superior)
Selecting Image (SI)	115 (Superior)
Completing Sentence (SE)	94 (Average)
Age	14 Year
RULE	IF ar (Superior) AND sa (Average) AND wa (Superior) AND ma (Highly Superior) THEN IQ (Superior)
Decision (D)	Superior

Fig 2. System Calculation Result

From Fig 2, it can be concluded that the results of the calculation process on the system are similar with the manual calculation results.

## C. Conclusive results on System Testing

The test was performed using 50 IST test result data. The IST test result data is compared with the results of the classification performed by the system.

TABLE 2. TEST DATA

Data	Classification by Psychologists	Classification by System	Information
1	Average	Average	Suitable
2	Above average	Above average	Suitable
3	Average	Average	Suitable
4	Average	Average	Suitable
5	Superior	Superior	Suitable
6	Average	Average	Suitable
7	Average	Average	Suitable
8	Average	Average	Suitable
9	Average	Unknow	Not Suitable
10	Superior	Superior	Suitable
11	Average	Average	Suitable
12	Above average	Unknow	Not Suitable
13	Very Superior	Unknow	Not Suitable
14	Average	Average	Suitable
15	Very Superior	Unknow	Not Suitable
16	Average	Superior	Not Suitable

17	Average	Unknow	Not Suitable
18	Average	Unknow	Not Suitable
19	Superior	Superior	Suitable
20	Average	Average	Suitable
21	Average	Average	Suitable
22	Superior	Superior	Suitable
23	Above average	Above average	Suitable
24	Average	Average	Suitable
25	Average	Average	Suitable
26	Average	Average	Suitable
27	Average	Average	Suitable
28	Above average	Above average	Suitable
29	Average	Average	Suitable
30	Average	Average	Suitable
31	Average	Average	Suitable
32	Average	Average	Suitable
33	Average	Average	Suitable
34	Superior	Superior	Suitable
35	Above average	Above average	Suitable
36	Above average	Above average	Suitable
37	Average	Average	Suitable
38	Above average	Above average	Suitable
39	Superior	Superior	Suitable
40	Superior	Superior	Suitable
41	Very Superior	Unknow	Not Suitable
42	Average	Average	Suitable
43	Very Superior	Very Superior	Suitable
44	Above average	Above average	Suitable
45	Average	Average	Suitable
46	Average	Average	Suitable
47	Superior	Superior	Suitable
48	Average	Average	Suitable
49	Average	Average	Suitable
50	Average	Average	Suitable

The following is a table of confusion matrix

TABLE 3. ACCURACY OF CONFIRMATION MATRIX

		Hasil Prediksi				
		Average	Above Average	Superior	Very Superior	Unknown
Original Class	Average	26	0	1	0	3
	Above Average	0	7	0	0	1
	Superior	0	0	8	0	0
	Very Superior	0	0	0	1	3
	Unknown	0	0	0	0	0

After the system classifies, then it calculates the value of its accuracy. The formula of accuracy is:

$$\begin{aligned}
 \text{Accuracy} &= \frac{\text{the amount of data predicted correctly}}{\text{number if predictions performed}} \\
 &= \frac{26+7+8+1}{26+7+8+1} \\
 &= \frac{26+1+3+7+1+8+1+3}{42} \\
 &= \frac{50}{42} \\
 &= 0.84
 \end{aligned}$$

The accuracy test data used in table 3 was 50 data, the IQ classification result by psychologist consisted of 30 "Average" IQ category, 8 "Above Average" IQ category, 8 "Superior" IQ category, and 4 "Very Superior" IQ category. The results of the classification performed by the system were as many as 50 data consisting of 26 "Average" IQ category, 7 "Above Average" IQ category, 9 "Superior" IQ category, 1 "Very Superior" IQ category and 7 "Unknown" IQ category. Then the number of classifications with the correct system is 42 classification of student IQs based on IST test result. Based on the accuracy test, the classification system of high school student IQs based on IST test results from 50 test data is 84%. As we all can see that C4.5 algorithm can be used as classification algorithm for student IQs test based on IST test results due to its high accuracy.

#### REFERENCES

- [1] Kusrini, Emha T. Lutfi, 2009, " Data Mining Algorithms ", Publisher ANDI, Yogyakarta.
- [2] Suryani, D., Labellapansa, A., & Marsela, E. (2018). *Accuracy of Algorithm C4.5 to Study Data Mining Against Selection of Contraception*. Paper presented at the Proceedings of the Second International Conference on the Future of ASEAN (ICoFA) 2017–Volume 2.
- [3] Azwar, Saifuddin., 2014, Pengantar Psikologi Inteligensi, PUSTAKA PELAJAR, Yogyakarta.
- [4] Kumolohadi, R., Suseno M. N., 2012, Intelligenz Struktur Test dan Standard Progressive Matrices (Dari Konsep Intelligensi yang Berbeda Menghasilkan Tingkat Intelligensi yang Sama), Jurnal Inovasi dan Kewirausahaan, Vol.1.
- [5] Lembaga Pengembangan Sarana Pengukuran dan Pendidikan Psikologi Universitas Indonesia (LPSP3 UI), 2014.
- [6] Prasetyo, E., 2014, Data Mining Mengolah Data Menjadi Informasi, ANDI, Yogyakarta.
- [7] Santosa, B., 2007, Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis, Graha Ilmu, Yogyakarta.
- [8] Hermawati, Fajar Astuti., 2013, Data Mining, ANDI, Yogyakarta.

# Data Mining Classification of Intelligence Quotient in High School Students

---

ORIGINALITY REPORT

---

**22%**

SIMILARITY INDEX

**21%**

INTERNET SOURCES

**18%**

PUBLICATIONS

**9%**

STUDENT PAPERS

---

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

---

6%

★ Des Suryani, Ause Labellapansa, Eka Marsela.  
"Chapter 95 Accuracy of Algorithm C4.5 to Study Data Mining Against Selection of Contraception",  
Springer Science and Business Media LLC, 2018  
Publication

---

Exclude quotes On

Exclude bibliography On

Exclude matches < 1%