International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 03, 2021

Psychological Evaluation Of College Students: A Data Mining Approach Using C5.0 Algorithm

Dr. D. Saravana priya[1], Dr. V. Sujitha[2], Dr. M. Umaselvi[3]

[1,2,3]*Department of CSE P. A. College of Engineering and Technology,Pollachi*

Email: [1]*sujithavpacet@gmail.com,* [2]*dspriyapacet@gmail.com,* [3]*umaselvipacet@gmail.com*

ABSTRACT: Students entering college have different characteristics, which can impact their performance in the classroom. The college conducts the Intelligence Structure Test to test the IQ of the students entering the campus. Till these days' classification process is done manually which takes long number of days. Data mining technique is used for classifying the required patterns from the data set. C5.0 algorithm is used for classification to minimize the time of psychologists in classifying data. This examination comprised of 10 standards including 350 information and brought about yield as four sorts of understudy's IQ order comprising of extremely unrivaled, predominant, better than expected and normal classifications. In light of the test outcomes arrangement of understudy IQs is acquired as 86%.

Keywords: Data mining, Intelligence structure test, Intelligence Quotient, C5.0 algorithm

1. INTRODUCTION

An instructive establishment of Pollachi, Coimbatore area (Autonomous Institution) has been directing mental tests utilizing IST as a brain research test in deciding the characterization of their understudy IQs. The IST test utilized is a Rudolf Amthaeur instrument which comprises of 10 subtests: age, finishing sentence (SE), word condition (WA), verbal relationship (A), character closeness (GE), mathematical capacity (RA), number arrangement (ZR), choosing picture (FA), 3D square drill (WU), and recollections (ME) and to decide the size of the IQ understudy grouping, they use David Wechsler's scale which comprises of 7 classes: profoundly predominant, prevalent, better than expected, normal, beneath normal, fringe and intellectually impeded. The grouping cycle of understudies IQs has been taken care of physically by analysts and took a normal of 15 days. This examination applies information mining which is a progression of cycles of getting information or example from informational collection utilizing technique for order calculation C5.0, which is helpful to rearrange and spare time for clinicians.

The development of data innovation is so much propelled today, causing the degree of exactness of an information is exceptionally required in regular daily existence. Every data that disseminates turns into something essential to decide on each choice in a specific circumstance. This prompts the arrangement of data to be broke down and summed up into an information on valuable information when a choice is made, ie

information mining.

Information mining is a lot of procedures to pick up information or examples from informational index utilizing a specific strategy or technique. The grouping strategy utilized is the Decision Tree, which is one of the simple order technique for human understanding. so that the inference process can be faster [1]. The structure resembles a tree, where the center point hints a property, the branch is the assessment of the quality and the leaf is the class[18]. The calculation utilized is C5.0 calculation. C5.0 calculation is utilized to frame littler choice trees. This calculation can likewise rearrange the information on the framework.

The examination results and modeling process using Decision Tree C5.0 algorithm conducted on the data of contraceptive use in Prabu Poly clinic can be made inference that the consequences of preparing information investigation of prophylactic use devices can frame information as choice tree models that change information into choice trees that speak to the principles. Exactness rate acquired from information preparing arrive at 93.15% that can be characterized as Excellent with the goal that this outcome can be a reference for birthing specialist in prompt patient in picking contraception [2].

2. LITERATURE REVIEW

A. Intelligence Quotient

(IQ) is a variety or the entirety of a person's ability to act in a particular explanation, think reasonably, and face the overall condition sufficiently. One of the most generally perceived ways used to convey high IQ level is to interpret the information test results into numbers that can be normal for a person's IQ level when reasonably diverged from a norm [3].

B. Intelligence Structure Test (IST)

IST test is one of the mental tests used to gauge singular knowledge [4]. IST comprises of nine subtests adding up to 176 things. The nine subtests in the IST are Satzerganzung (SE) finishing the sentence, Wortauswahl (WA) word condition, Analogien (A) verbal relationship, Gemeinsamkeiten (GE) character closeness, Rechhenaufgaben (RA) mathematical capacity, Zahlenreihen (ZR) number arrangement, Figurenauswahl FA) choosing picture, Wurfelaufgaben (WU) shape drills, and Merkaufgaben (ME) recollections [5].

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

IST	IQ	Category
≥ 120	≥ 131	Highly intelligent
113 – 119	121 – 130	Intelligent
107 – 112	111 – 120	Good
93 – 106	91 – 110	Not bad
87 – 92	88 – 90	Below not bad
80 – 86	81 – 87	Okay
≤ 79	≤ 70	Mentally retarded

C. Data Mining

Data mining is a chain of tasks to gain expertise or patterns from data set [6]. Data Mining resolve the problem by investigating the information which is as of now in the

database. Information Mining, frequently called information disclosure in databases (KDD), is a movement that incorporates information assortment, utilization of recorded information to discover examples of consistency and relationship designs in enormous informational indexes [7]. The yield of information mining can be utilized to improve future dynamic.

D. Classification

The grouping method is a methodical way to deal with build an arrangement model of an information informational index [8]. Characterization is the learning cycle of an objective or an objective capacity that maps each arrangement of x ascribes to one of the recently characterized y names. The objective capacity is likewise called the grouping model [9][19].

E. Decision Tree

This technique utilized tree structure to fabricate the characterization models. It partitions a dataset into littler subsets [10]. Leaf hub speaks to a choice. In view of highlight estimations of occurrences, the choice trees group the cases. Every hub speaks to a component in an example in a choice tree which is to be arranged, and each branch speaks to a worth [11]. Order of Instances begins from the root hub and arranged dependent on their component esteems. Unmitigated and mathematical information can be taken care of by choice trees. [2]

F. C5.0 Algorithm

Feature Selection:
Information with amazingly high dimensionality has introduced genuine difficulties to existing learning techniques [6]. Due to enormous number of highlights, a learning model create the outcome which is over-teed, so the presentation is debased [12]. For diminishing dimensionality, the element choice is a generally utilized method [20]. It picks a little subset of the pertinent highlights from the real set which ordinarily gives better learning execution and model interpretability. Utilizing Feature choice method, the computational expense of learning can likewise decrease.

Reduced Error Pruning:
In order models, it is regular practice to dispose of parts of the model that portray misleading impacts in the preparation test as opposed to genuine highlights [13]. REP is a procedure that eliminates areas of the tree to lessen the size of choice trees[17]. It eliminates the aspect of the tree which gives less capacity to the order of examples. The Pruning method decreased intricacy and gives better precision by lessening over fitting issue.

Algorithm:

- Create a root node
- Check the base case
- **By using Genetic Search Apply Feature Selection technique**

bestTree = a decision tree is constructed by using training data

- **Apply Cross validation technique**

1. Split all training data into N disjoint subsets, R = R_1, R_2, ..., R_N
2. For each j = 1, ..., N do

Test set = R_j
Training set = $R - R_j$
Enumerate the decision tree by using Training set
Adjudge the performance accuracy X_j by using the Test set
3. Compute the N-fold cross-validation technique to estimate the performance
= $(X_1 + X_2 + ... + X_N)/N$

Finally apply the reduced error pruning technique
Find the attribute with the highest info gain (A_Best) Classification:
For each tj ϵ D, apply the DT to determine its class.

G. Accuracy

An arrangement framework is required to group all informational indexes effectively[14]. In any case, it can't be rejected that the presentation of a framework can't work 100% effectively [6]. In this manner, an order framework ought to likewise be estimated for execution. By and large estimating arrangement execution utilizes disarray framework. The disarray grid is a table that records the aftereffects of grouping work.

Accuracy = the amount of correctly predicted data
Number of predictions performed

3. RESEARCH METHODOLOGY

A. Data Collecting

The information assortment methods required in the characterization of IQs, gotten from meetings and writing study.

- Interviews, led to gather data that would be helpful in arranging understudy IQs. The meeting was directed on therapists who arrange brain research tests in instructive organization, Pollachi.
- Literature study, gathering information via looking and concentrating from different sources identified with the issues concentrated in the arrangement of this proposition, both from books, logical diaries and from different readings that can be represented.

4. RESULTS AND DISCUSSIONS

From the clarification of the investigation and the plan that has been done in the past section related, the following stage is to actualize and test the framework by testing discovery which is a test that centers around the practical necessities of the product[15]. In this way, empowering programming architects to get a lot of information conditions that completely use all useful prerequisites of a program. At that point by testing end clients to get great outcomes prompting the last target of the underlying arrangement of the framework[16].

A. Testing using C5.0 Algorithm

By using the C5.0 algorithm testing is performed for 350 student data, through which the training data is classifies IQ classes based on 4 categories namely, highly intelligent (> 131), intelligent (121 - 130), good (111 - 120), and not bad (91 - 110). There are 10 criteria used for IQ classification with C5.0 algorithm, with details as follow:

- WA criteria (word equation), consisting of 6 categories that are highly intelligent (> 131), intelligent (121 - 130), good (111 - 120), not bad (91 - 110), below not bad (88- 90), and okay (81 - 87).
- AN Criteria (verbal analogy), consisting of 6 categories such highly intelligent (> 131), intelligent (121 - 130), good (111 - 120), not bad (91 - 110), below not bad (88- 90), and okay (81 - 87).
- ME criteria (memory), consisting of 6 categories which are highly intelligent (> 131), intelligent (121 - 130), good (111 - 120), not bad (91 - 110), below not bad (88- 90), and okay (81 - 87).
- GE criteria (the same nature), consisting of 6 categories with highly intelligent (> 131), intelligent (121 - 130), good (111 - 120), not bad (91 - 110), below not bad (88- 90), and okay (81 - 87).
- RA criterion, consisting of 6 categories, they are highly intelligent (> 131), intelligent (121 - 130), good (111 - 120), not bad (91 - 110), below not bad (88- 90), and okay (81 - 87).
- WU criteria (cube exercises), consisting of 6 categories: highly intelligent (> 131), intelligent (121 - 130), good (111 - 120), not bad (91 - 110), below not bad (88- 90), and okay (81 - 87).
- ZR criteria (series of numbers), consisting of 6 categories: highly intelligent (> 131), intelligent (121 - 130), good (111 - 120), not bad (91 - 110), below not bad (88- 90), and okay (81 - 87).
- FA Criteria (selecting images), consisting of 6 categories: highly intelligent (> 131), intelligent (121 - 130), good (111 - 120), not bad (91 - 110), below not bad (88- 90), and okay (81 - 87).
- SE criteria (complete sentence), consists of 6 categories: highly intelligent (> 131), intelligent (121 - 130), good (111 - 120), not bad (91 - 110), below not bad (88- 90), and okay (81 - 87).
- Age criteria, consisting of 4 categories ie age 18, age 19, age 20, and age 21.

1. Entropy Esteem Calculation

The first step of the C5.0 calculation is to discover the entrophy esteem. First is to decide the complete entropy esteem for the situation. The entropy relies on every rule. Next is to ascertain the addition esteem for every basis.

2. Gain Value Calculation

After all entropy figuring on every rule are finished, at that point the increase esteem is determined. The most elevated addition estimation of 0.3503 is made as the principal root hub. For the following root hub, count will be done again until all framed examples are met.

3. Determining the Root

From the figuring to discover the estimation of entropy and increase esteem, it tends to be distinguished that quality with most noteworthy addition is ZR which is equivalent to 0.3575. In this manner, ZR turns into the root hub. The information will at that point be separated into 6 gatherings: "prevalent", "predominant", "better than expected", "normal", "underneath normal", and "fringe". The accompanying choice tree is shaped for hub 1:

B. System Testing

1. Input value of each criterion

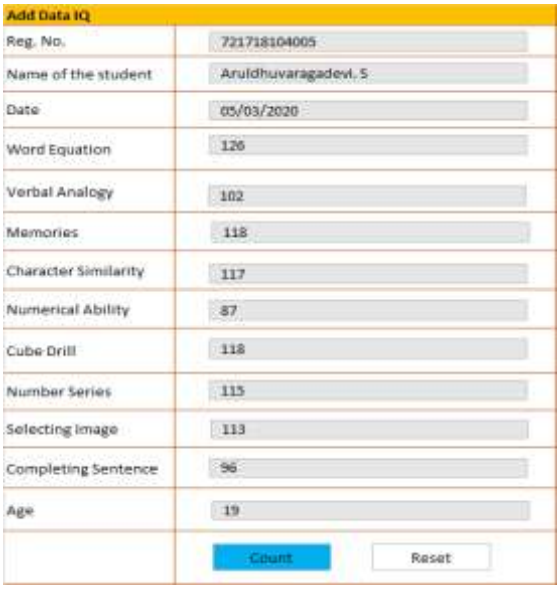

Fig.1 Criterion value input

From Fig 1, it very well may be clarified that the contribution of basis esteem is equivalent to the standard in manual test.

2. College student's IQ Classification Output

IST Test Result	
Reg. No.	721718104005
Name of the student	Aruldhuvaragadevi .S
Date	5/3/2020
Word Equation	126 (Highly Superior)
Verbal Analogy	117 (Superior)
Memories	127 (Highly Superior)
Character Similarity	117 (Superior)
Numerical Ability	106 (Average)
Cube Drill	118 (Superior)
Number Series	115 (Superior)
Selecting Image	113 (Superior)
Completing Sentence	96 (Average)
Age	19 Year
Rule	IF VA(superior) AND NA (Average) AND CD(Superior) AND ME(HighlySuperior) THEN IQ(Superior)
Decision	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 utilizing 50 IST test result information. The IST test result information is contrasted and the consequences of the grouping performed by the framework.

TABLE 2. TEST DATA

Data	Classification by Psychologists	Classification by System	Information
1	Not bad	Not bad	Suitable
2	Good	Good	Suitable
3	Not bad	Not bad	Suitable
4	Not bad	Not bad	Suitable
5	Intelligent	Intelligent	Suitable
6	Not bad	Not bad	Suitable
7	Not bad	Not bad	Suitable
8	Not bad	Not bad	Suitable
9	Not bad	Not bad	Suitable
10	Intelligent	Intelligent	Suitable
11	Not bad	Not bad	Suitable
12	Good	Unknown	Not Suitable
13	Highly intelligent	Unknown	Not Suitable
14	Not bad	Not bad	Suitable
15	Highly intelligent	Unknow	Not Suitable
16	Not bad	Intelligent	Not Suitable
17	Not bad	Unknow	Not Suitable
18	Not bad	Unknow	Not Suitable
19	Intelligent	Intelligent	Suitable
20	Not bad	Not bad	Suitable

21	Not bad	Not bad	Suitable
22	Intelligent	Intelligent	Suitable
23	Good	Good	Suitable
24	Not bad	Not bad	Suitable
25	Not bad	Not bad	Suitable
26	Not bad	Not bad	Suitable
27	Not bad	Not bad	Suitable
28	Good	Good	Suitable
29	Not bad	Not bad	Suitable
30	Not bad	Not bad	Suitable
31	Not bad	Not bad	Suitable
32	Not bad	Not bad	Suitable
33	Not bad	Not bad	Suitable
34	Intelligent	Intelligent	Suitable
35	Good	Good	Suitable
36	Good	Good	Suitable
37	Not bad	Not bad	Suitable
38	Good	Good	Suitable
39	Intelligent	Intelligent	Suitable
40	Intelligent	Intelligent	Suitable
41	Highly intelligent	Unknow	Not Suitable
42	Not bad	Not bad	Suitable
43	Highly intelligent	Highly intelligent	Suitable
44	Good	Good	Suitable
45	Not bad	Not bad	Suitable
46	Not bad	Not bad	Suitable
47	Intelligent	Intelligent	Suitable
48	Not bad	Not bad	Suitable
49	Not bad	Not bad	Suitable
50	Not bad	Not bad	Suitable

The following table betrays the confusion matrix:

TABLE 3. ACCURACY OF CONFIRMATION MATRIX

		Hasil Prediksi				
		Not bad	**Good**	**Intelligent**	**Highly intelligent**	**Unknown**
Original Class	Not bad	26	0	1	0	3
	Above Not bad	0	7	0	0	1
	Intelligent	0	0	9	0	0
	Very Intelligent	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:

$$\text{Accuracy} = \frac{\text{the amount of correctly data predicted}}{\text{Number of predictions performed}}$$

$$= \frac{26+7+9+1}{}$$

$$= \frac{26+1+3+7+1+8+1+3}{50}$$
$$= \frac{43}{50}$$
$$= 0.86$$

The precision test information utilized in table 3 was 50 information, the IQ arrangement result by therapist comprised of 30 "Normal" IQ class, 8 "Better than expected" IQ classification, 8 "Prevalent" IQ class, and 4 "Exceptionally Intelligent" IQ classification. The consequences of the order performed by the framework were upwards of 50 information comprising of 26 "Normal" IQ classification, 7 "Better than expected" IQ classification, 9 "Predominant" IQ class, 1 "Extremely Intelligent" IQ class and 7 "Obscure" IQ class. At that point the quantity of orders with the right framework is 42 characterization of understudy IQs dependent on IST test result. In light of the precision test, the order arrangement of undergrad IQs dependent on IST test results from 50 test information is 86%. As we as a whole can see that C5.0 calculation can be utilized as arrangement calculation for understudy IQs test dependent on IST test results because of its high exactness.

5. 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 Strutur
[5] Test dan Standard Progressive Matrices (Dari Konsep Inteligensi yang Berbeda Menghasilkan Tingkat Inteligensi yang Sama), Jurnal Inovasi dan Kewirausahaan, Vol.1.
[6] Lembaga Pengembangan Sarana Pengukuran dan Pendidikan Psikologi Universitas Indonesia (LPSP3 UI), 2014.
[7] Prasetyo, E., 2014, Data Mining Mengolah Data Menjadi Informasi, ANDI, Yogyakarta.
[8] Santosa, B., 2007, Data Mining Teknik Pemanfaatan Data
[9] Untuk Keperluan Bisnis, Graha Ilmu, Yogyakarta.
[10] Hermawati, Fajar Astuti., 2013, Data Mining, ANDI, Yogyakarta.
[11] Krishnan, P, Duttagupta, S, Achuthan, K. SDN/NFV security framework for fog-to-things computing infrastructure. Softw: Pract Exper. 2020; 50: 757–800. https://doi.org/10.1002/spe.2761
[12] Krishnan P., Achuthan K. (2019) CloudSDN: Enabling SDN Framework for Security and Threat Analytics in Cloud Networks. In: Kumar N., Venkatesha Prasad R. (eds) Ubiquitous Communications and Network Computing. UBICNET 2019. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 276. Springer, Cham. https://doi.org/10.1007/978-3-030-20615-4_12
[13] Krishnan P., Achuthan K. (2019) Managing Network Functions in Stateful Application Aware SDN. In: Thampi S., Madria S., Wang G., Rawat D., Alcaraz Calero J. (eds) Security in Computing and Communications. SSCC 2018. Communications in Computer and Information Science, vol 969. Springer,

Singapore. https://doi.org/10.1007/978-981-13-5826-5_7

[14] Krishnan, P., Duttagupta, S. & Achuthan, K. SDNFV Based Threat Monitoring and Security Framework for Multi-Access Edge Computing Infrastructure. Mobile Netw Appl 24, 1896–1923 (2019). https://doi.org/10.1007/s11036-019-01389-2

[15] Krishnan, P., Duttagupta, S., & Achuthan, K. (2019). VARMAN: Multi-plane security framework for software defined networks. Comput. Commun., 148, 215-239. https://doi.org/10.1016/j.comcom.2019.09.014

[16] K. Raghunath and P. Krishnan, "Towards A Secure SDN Architecture," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bangalore, 2018, pp. 1-7, doi: 10.1109/ICCCNT.2018.8494043.

[17] Krishnan P., Najeem J.S., Achuthan K. (2018) SDN Framework for Securing IoT Networks. In: Kumar N., Thakre A. (eds) Ubiquitous Communications and Network Computing. UBICNET 2017. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 218. Springer, Cham. https://doi.org/10.1007/978-3-319-73423-1_11

[18] Krishnan, P., Jose, P.G., Jain, K., Achuthan, K., & Buyya, R. (2019). SDN Enabled QoE and Security Framework for Multimedia Applications in 5G Networks. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM).DOI:https://doi.org/10.1145/3377390.

[19] Lakshmanaprabu S.K, Mohanty.S. N,Sheeba Rani,, Sujatha Krishnamoorthy, Uthayakumar ,Sankar(2019"Online clinical decision support system using optimal deep neural networks" Volume 81, August 2019, 105487, Applied Soft Computing, Elsevier

[20] Rajadevi R,Johnpaul P,Sathish Kumar S,K Venkatachalam, Kowsigam M, A hybrid optimization memethic algorithm to solve cloud scheduling problems, Journal of Green Engineering, 2020, 10(3), pp.986-997