

Extracting Knowledge Concerning Digital Literacy by using a Predictive Model in Data Mining

Soheila Sadeghiyan
Department
of instructional technology
Islamic azad university-south
Tehran branch
Tehran, Iran

Dariush Noroozi, Ph.D
Department
of instructional technology
Allameh Tabatabaei University
Tehran, Iran

Marjan Maadi
Department of IT Engineering
Graduate University of
Advanced Technology
Kerman, Iran

ABSTRACT

In this paper, it is intend to implement four different algorithms in data mining (C5.0, C&R tree, QUEST ,CHAID) on the data related to the level of familiarity with computers and Internet. The goal of this paper is prediction of the effect of computer literacy and internet literacy on Digital literacy competence of individuals. To determine best predictive model, the accuracy of generated model is obtained. The prediction of Digital literacy competence shows the ability and skills of individuals to acquisition of new technologies, because using new technologies leads to reach performance improvement and higher motivations.

Model of C&R tree with 73.2%, has had the best accuracy. The data in this study are in the form of questionnaire and were taken from undergraduate students of Islamic Azad University of south Tehran branch. This paper uses master's thesis data with the titles of the relationship between digital literacy and academic performance and progress motivation of students of graduate school of literature and Foreign languages of Islamic Azad university of south Tehran.

General Terms

Data mining, clustering, Digital literacy.

Keywords

Data mining, clustering, Digital literacy competence, internet literacy.

1. INTRODUCTION

Nowadays, with considering changes and developments, Vast amount of information and the production of unimaginable level of knowledge in different fields of science, everybody requires to learn skills and capabilities beyond the traditional literacy to keep up with these changes and developments, particularly in use of information the key to succeed in this era, is being equipped with these capabilities so that a person can recognize his information needs, and then choose the most appropriate source of information. He needs to locate and evaluate the required information by critical thinking .Thus, people living in the 21st century, for reconciling with these developments, require learning and applying new skills which can be used for the conception of digital literacy. This includes needed skills for progress in digital era. This conception was used for a set of skills and competence for users in digital environments. It consists of a set of skills for work, education and social life.

This literacy can be introduced through digital instruments and includes awareness, insight and ability of individuals in using them, availability, management, integration, evaluation, analysis and combining of digital sources, producing new

technologies, creating media status and communication with others [1].

Digital literacy skills have great effects on learning process in life from availability of information and applying sources .So, this literacy has a major role based on obtained skills and knowledge. A person with digital literacy can compile new technologies and new lingual symbols rapidly.

A vast amount of information can be released in digital environments today. There is a necessity for coordination with rapid growing of knowledge using of electronic information students with proper insight of instrument and search facilities. Digital literacy is an important element in this field. Changing of pressing technologies to digital brought about a change in the way students study, because this new generation is growing in digital world [2].

Digital literacy is an important title for youth in developing digital world. This world is based on digital technologies including computer, communicative networks, software, electronic sources and other communication technologies. This literacy makes familiar young people with economic and social progresses and we cannot ignore the cultural and social operation of digital literacy that enables us to interact with this world of digital technologies. Regarding the importance of digital literacy new technologies,

Digital literacy has positive effects on academic education of students, their motivation and better management of policies made in the field. In future, on the knowledge- based societies can survive which have members with required skills for their life time.

2. THEORETICAL DISCUSSION

Guilster, for the first time, has introduced this conception with such a title. This conception is beyond of a set of particular skills. The basic elements of this conception are: analytical and critical issues. In other words, this literacy is a kind of insight about learning of various sets of skills for cognition capabilities like critical thinking and analysis. A decade after introducing digital literacy, development of technologies like Google and social networks, made this conception of high validity[3].

Gulster [3] has showed not only a list of skills in digital literacy, but he has explained that this is a kind of ability for understanding and using data from different sources, it is called literacy in digital era. As a result, he points to traditional literacy as reading, writing and interaction with information by using new forms of technologies. Guilster says that digital literacy is the ability for understanding information and evaluating it with different forms, where evaluation and

interpretation of information is Important. He emphasizes those people who have access to internet without evaluating sources, cannot understand them in proper structure. This problem has two forms: First, data forms on the internet are not only restricted to texts and include multi-media files. Second, methods for accession to information are different with traditional methods.

Guilster believes that critical thinking skills are important for evaluating information on internet. It is one of the elements of digital literacy [4]. Digital literacy is a structure that consists of three levels: First level is digital ability of users for necessary duties in digital life. This level for all of users' activities is an important measure. Second level includes some digital users with digital knowledge in a particular field and important processes in society. Third level is time-based level. Digital applications are developing and innovating. Martin called this stage as digital changing. There are no direct sequences between stages and digital literacy is a growing process with technology development. In fact, digital literacy includes proper using of digital capabilities in different situation. Digital literacy development is one of dimension of progress. It must be meaningful and indicates routine life. Digital literacy is not a goal, but a capability .It causes growing people in social life. Digital literacy has three dimensions, technical, cognitive and Psychological- social. Technical dimension means Technical skills and performance of it, learning and routine activities. Cognitive dimension is with ability to think critically in research, evaluation and creating digital information. Social – Psychological dimension is the region between social regions and use of internet for communication, social relationship and learning.

3. RESEARCH BACKGROUND

Marinak says in a research known as "*Surveying digital literacy use and motivation in elementary students*" that, new technologies are changing the essence of literacy and knowledge [5]. Researchers believe that the structure of digital literacy could improve the interest by making some opportunities to solve the problem in the real world. In this research, more interest and motivation to digital literacy is seen among the elementary school students in comparison with traditional literacy (reading and writing). In a study, among 93 high school students, an obvious statistical difference in value about digital literacy was identified between those who prefer reading printed texts and those who are interested in using computer [6].

Fombona et al. investigated in a research known as "*Digital Literacy and adult's learners' perception*" that how electronic instruction can help the adults to pursue their education and respond to the challenges of today society. So improving and developing of digital literacy has an important role in this area. The results of these studies prove that, students have a positive perspective toward using internet instrument. In this research the reason of selecting students is considered the need for improvement of academic level [6].

Anderson in his research known as " use of internet among students" which was done in universities of china, got to some conclusions which showed that the use of internet has improved the educational experience of students in universities of china, expanded the access to education and educational materials, improved the quality of education and educational curriculum, and increased creativity in students [7].

4. DATA COLLECTION INSTRUMENT

In this research, the Data collection instrument is questionnaire that is drawn up in [8]. The questions of the questionnaire are formulated on the basis of five processional skills belong Jules and Toman. Although they formulated these skills labeled as media literacy, but due to the fact that they include the whole skills needed in digital literacy this structure of skills is use Clementine 12.0 to reach the desired goal of a predictive model [9].

5. RESEARCH METHOD

The existing set of digital literacy data are collected from distribution of questionnaire among B.A students of Islamic Azad university_ south Tehran branch. This set of data includes 225 samples and 26 traits and does not have any lost amount. 24 traits of sample data of two branches of the questionnaire questions include computer literacy questions (basis skills of using computer_ 7 questions) and internet literacy questions (basis skills of using_ 17 questions). The other 2 traits are the age and gender of the question responders.

The algorithm of classification is implemented to reach to a predictive model in order to estimating the digital literacy of individuals. This goal is obtained by measuring the level of computer literacy and internet literacy of people. To do this, a variable is introduced to calculate the digital literacy of people. Then the existing data will be organized in ascending order, based on Computer literacy fields and internet literacy fields. In the text phase, the sample data is divided to two subcategories (70 percent for instruction and 30 percent for experiment) and at last the algorithms of C5.0, C& R Tree, QUEST, and CHAID and decision tree that are the algorithms of different rules deduction in Clementine 12.0 are applied on the data [10]. Decision tree or a collection of rules to describe distinct part of the related data, is branched from the mentioned algorithms.

These algorithms make the decision tree by recursive division of data to the subcategories that are dependent to conclusion and are defined by predictive fields.

6. EXPERIMENTAL AND RESULTS

Results show that the validity of models created by implementation of the algorithms of C5.0, C & R tree, CHAID, and QUEST on the collection of instructional data and test, are the values on Table 1.

Table 1. the extent of model's validity on the instructional and test data

Algorithm	Validity	
	Instruction data	Test data
C5.0	65.36 percent	37.5 percent
C & R Tree	73.2 percent	44.44 percent
CHAID	69.93 percent	47.22 percent
QUEST	63.4 percent	40.28 percent

To specify the best anticipant model of the amount of digital literacy, It is a need to study the extent of validity of produced models. The extent of validity of a classification method on the instructional data is the percent of observation of the instructional collection that are correctly classified by the used method. In the system of pattern recognition, this essential

index is called the rate of recognition that is the indicator of the quality of recognizing the samples of different categories. The value of validity is calculated by formula (1).

$$\text{Formula (1) Validity} = (\text{TP} + \text{TN}) / (\text{pos} + \text{Neg})$$

TP: the number of data labeled positive

TN: the number of data labeled negative

Pos: the whole number of positive data

Neg: the whole number of negative data (17)

The highest validity average (58.82) on the instructional and test data in the algorithm C & R Tree is the indicator of better function of this algorithm on the collection of reviewing data. The more is the validity of the model, the more compatibility of the sample data with the decision tree will be, and more accurate rules or more accurate knowledge could be elicited. So, we show the obtained rules of the best developed decision tree (decision tree, model c and R Tree) in Table 2.

Table 2: a sample of the rules made by the decision tree C& R Tree

Number	Rules
1	If $24.5 \geq \text{internet literacy} \geq 18.5$ and computer literacy ≤ 11.5 and $21.5 < \text{age} \leq 28$ then the number digital literacy is 26.
2	If $18.5 \leq \text{internet literacy} \leq 24.5$ and computer literacy ≤ 11.5 and $\text{age} < 21.5$ then the number of digital literacy is 27.
3	If $18.5 \leq \text{internet literacy} \leq 24.5$ and computer literacy ≤ 11.5 and $\text{age} > 28$ then the number of digital literacy is 27.
4	If $18.5 < \text{internet literacy} \leq 22.5$ and computer literacy ≤ 11.5 and $\text{age} > 36.5$ then the number of digital literacy is 28.
5	If $18.5 < \text{internet literacy} \leq 22.5$ and computer literacy ≤ 11.5 and $\text{age} \leq 36.5$ then the number of digital literacy is 30.
6	If $22.5 < \text{internet literacy} \leq 24.5$ and computer literacy ≤ 11.5 and $\text{age} > 24$ then the number of digital literacy is 34.
7	If $22.5 < \text{internet literacy} \leq 24.5$ and computer literacy ≤ 11.5 and $\text{age} < 24$ then the number of digital literacy is 35.
8	If $21.5 < \text{internet literacy} < 24.5$ and computer literacy ≤ 11.5 and $\text{age} > 20.5$ then the number of digital literacy is 37.
9	If $21.5 < \text{internet literacy} \leq 24.5$ and computer literacy > 11.5 and $\text{age} \leq 20.5$ then the number of digital literacy is 38.
10	If $28.5 < \text{internet literacy} \leq 30.5$ and $\text{age} \leq 22.5$ then the number of digital literacy is 45.
11	If $28.5 < \text{internet literacy} \leq 30.5$ and $\text{age} > 22.5$ then the number of digital literacy is 46.

To incorporate digital literacy in academic studies, there have been used mostly the same techniques in the related researches. The data, in these researches, are collected by using questionnaires or taking some tests in digital environments like internet, computer, etc . . . But there are two weak points about these techniques that effect the validity of the researches negative : firstly , they don't put the extracted data in an organizational and accurate model to make a thorough conclusion and secondly of the individual whom the study is done among .unlike these researches ,the present paper makes a predictive model known as decision tree by using the algorithms of C5.0, C&R tree, QUEST and CHAID that leads to accurate amount .Another aspect that this paper considers is the two factors of ages and gender of the individuals . by considering all of these factors and aspects , the conclusion is accompanied with high validity.

7. DISCUSSION

According to table 2, the results of the present research show that the extent of internet literacy has a great effect on digital literacy in a way that the results gained by implementing the Bayesian network indicate the important ratio of internet literacy 0.6, on the reviewing data. Also the importance ratio of computer literacy in digital literacy is 0.3 according to the results of Bayesian network. So in digital literacy, the extent of internet literacy is twice more important than computer literacy. The rules 1 to 9 in table 2 prove this subject matter. In the same conditions, regarding the effect of computer literacy and internet literacy, the age of user influence the extent of digital literacy. The young age shows that the extent of digital literacy is high (according to rules 4 to 9). So another effectual factor in the extent of digital literacy is the age of user. The younger the user, the higher is his or her extent of digital literacy. The cause of this function perhaps is the more interest in using to increase skills and learning among users. Also the higher ability of younger user could be an effective factor.

Regarding the results of the research, the existence of an organized attitude in order to predicting the extent of digital literacy of user to help the developers of digital literacy with the purpose of correct recognition of the extent of user's ability and skills, is necessary. One the improvable parts of this matter is, anticipating the extent of digital literacy of the user new technologies in organizations, with the goal of improving organizational functions, expanding new technologies ,and making backup systems to help the developer of digital literacy.

8. CONCLUSION

In this research classification algorithms, C5.0, C&R Tree, QUSET and CHAID is applied on the digital literacy data and a decision tree was made to predict the extent of digital literacy of individual . The prediction of the extent of digital literacy is done with the purpose of obtaining some cases like helping the developers of digital literacy, to identify the extent of user's abilities and skills correctly in the use of new technologies and increasing the internet of the developers of digital literacy to introduce new worldly technologies .Results show that the decision tree made by the algorithm, C&R Tree is more appropriate because it has higher validity. So the elicited rules from this model were used to explore new knowledge. The obtained rules show that internet literacy is twice more effectual than computer literacy regarding the extent of digital literacy. Also in the same condition, the factor of user's age could be effectual on digital literacy.

9. REFERENCES

- [1] Martin, A.2005. .DigEuLit- aEuropean framework for Digital literacy: a progress report. *Journal of ELiteracy*, 130-136.
- [2] Future lab 2010. *Digital Literacy across the curriculum: A futurelab handbook*. Retrieved from [www Futurelab.org.uk](http://www.Futurelab.org.uk)
- [3] Gilster, Paul 1997, *Digital Literacy*, London: John Wiley & Sons, Inc.
- [4] Martin, A. 2006 .A European framework for Digital Literacy .*Nordic Journal of Digital Literacy*, 2,151- 160.
- [5] Marinak, B., Evan, O (2013). Surveying digital literacy use and motivation in elementary school students, *BasicResearch Journal of Education Research and Review* ISSN 2315-6872 Vol .2 (5) pp. 81.88
- [6] Fombona, J., Goulao, M. 2012. Digital Literacy and adults learners' perception: The case of a second chance to University Procardia-Social and Behavioral Science 46, 350-355.
- [7] Anderson, S. 2006. The evaluation of higher education in china: Storied experiences of Internet usage by student and faculty. A Dissertation presented to the faculty of the College of Education University of Houston. Available from: <http://www.proquest .com> [Accessed 11 July 2009].
- [8] Ahmadi, M. 2010. Tehran University students studying digital literacy and Provide a framework for the promotion of digital literacy, social science master's thesis, Tehran University.
- [9] Modeling techniques in Clementine. Chapter11 Available from:
URL:<https://fhss.byu.edu/SPSS%20Modeler/Chapter%2011.pdf>
- [10] Rule Induction. Chapter12; Available from:
URL:<https://fhss.byu.edu/SPSS%20Modeler/Chapter%2012.pdf>