Studying Factors Affecting English Language Learning in University Students by Data Mining

Mona Naghdeforoshha[1] and Fahimeh Taherkhani[2]

[1, 2] Department of computer, Faculty of science, Takestan branch, Islamic Azad University, Takestan, Iran.

ABSTRACT

Language is the most developed mental tool which is used for making relation between humans and human societies. Transmission of complicated mental concepts is only possible through written and oral language in the simplest and the most complete way. The main objective of this project is finding a model and approach for improving and increasing the university students' success in English language. The performed researches determine that in present research, for the first time in Iran, Data Mining is used to fulfill this objective. To do so 100 informational forms was prepared from students of Mechanics Engineering and Agricultural Engineering of Islamic Azad University, Takestan Branch. 70% of data was used for analysis and 30% for model testing. In this article, Classification algorithm and Clustering are used for analyzing data by Data Mining. In clustering algorithm, data are placed in their appropriate clusters considering the similarities of data. The quality of these clusters is not clear in advance. The obtained results in this method are based on clusters' analysis. In classification algorithm the quality of groups is clear in advance. Data are placed in their appropriate group and groups are studied.

KEYWORDS: English language learning, Data mining, Classification, Clustering.

1. INTRODUCTION

In the age of technology and communications and ever increasing electronic training, there is no excuse for not learning any more. Today learning English language is an undeniable necessity; because English language is the language of science and learning different fields in advanced levels and producing science take place in English language. Nobody can deny language classes or efforts of caring professors and teachers of English languages. But you can learn language in any condition. But what kind of university students are successful in learning English language?

There are numerous factors engaged in success of students in learning English language including level of self-confidence, age of learning English language, place of learning English language, student's aim by learning this language and … [1-3]. From the reasons of using data mining we can mention the ability of data mining in studying these factors. In addition we can extract practical factors and models amongst data through data mining.

Methods used in this article are clustering and classification. Clustering is a kind of classification for objects according to their level of similarity but the number of groups is not determined beforehand. In clustering any identity is compared with other object. In case of similarity they will be placed in one group. In order to classify the data we can use artificial nervous networks, decision tree, and biz. In this article for classifying data, algorithm of decision tree was used. In decision tree, samples are transferred to the various tree branches according to the quantities of their properties [4-5].

2. Data Mining

Data mining is a science for extracting information and discovering hidden patterns from a multitude of great data bases. Using this science provide us with patterns by using which we can answer challenges established about considered subject. Data mining stages are as follows:
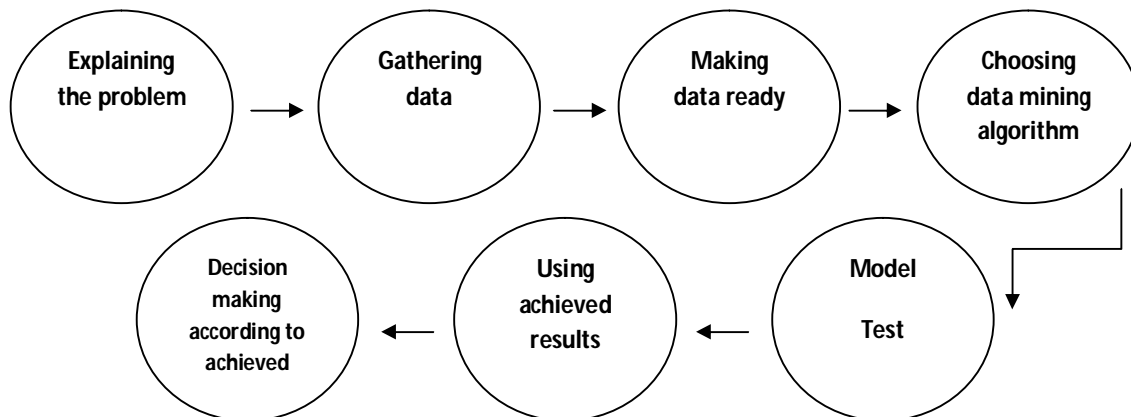

Fig.1. Data mining stages

Corresponding Author: Mona Naghdeforoshha, Department of computer, Faculty of science, Takestan branch, Islamic Azad University, Takestan, Iran. e-mail address: m_naghdi@yahoo.com, (phone: +98-0282-527-0130)

3. Method of Performance

Data set of this section includes informational form prepared from 100 university students of Mechanics Engineering and Agricultural Engineering of Islamic Azad University, Takistan Branch. Gender of 40 people is female and 60 others are male. Majority of them are unemployed and single. Average GTA of people under study was 16 and majority of them had average age of 20-23. Data qualities are briefly as follows:

Table 1. Data chart

Gender	Number	Single	Married	Employed	Unemployed
Woman	40	27	13	9	31
Man	60	46	14	21	39

We entered information of these forms in Excel software. The students whose total score of fluency was lower than 5 out of 10 were considered as unsuccessful students in learning English language; score 6 as mead students and 7 and more as successful students in learning English language. Then a data base was prepared. This data base was studied and processed by WEKA statistical software.

In this article data are studied through clustering method and using K-means algorithm. In this algorithm, K is the number of clusters that is taken from input. Present objects are divided into K clusters according to their similarity to mead of objects of clusters. For determining the quantity of similarity, different methods are used some of which are used in data mining and some do not have vast utilization. Also, data were studied using algorithm of decision tree. Studying data by this method and placement of data in their appropriate group, a tree was achieved. Leaves of the tree that has less than five objects has no statistical value. These kinds of leaves were omitted. After studying these data through these algorithms, appropriate models were achieved.

4. Summary of Performance

For controlling the accuracy of obtained results through software, we separated 10 forms and in fact we just studied 90 forms. Finally we had a data base including 90 records each having 21 fields.

In data studying by clustering and K-means algorithm, omitting tree fields of Marital Status, Employment Status, and Studying Field because of being the same and ineffective and increase of number of clusters to 3 we reached to desirable results. Then, using classification method and algorithm of decision tree, we processed data. In this method we omitted some of the fields that cause tree to be scattered and their importance seems less were omitted and desirable tree was achieved. Finally data were classified in three groups G (successful), M (mead), and B (unsuccessful).

5. Comparison of Two Methods Applied on Data

Percentage of model errors was obtained through 10 records that were put aside from the beginning for test model. The obtained model through clustering method and K-means algorithm had 30% error. Also, the obtained model through classification and decision tree method had 10% error. This illustrates that the obtained pattern through classification and decision tree method is more appropriate.

6. Obtained Results

After studying data through mentioned algorithms some results were obtained which show us the qualities of successful, mean, and unsuccessful students in learning English language. Obtained results can be used for analyzing the amount of success in learning. Obtained results are explained below.

6.1. Results of Data Processing by Clustering and K-means Algorithm Method

Using this algorithm on data, they were divided in 3 clusters. After changing number of clusters to three clusters, all results that we had in studying data appeared and following figure results were obtained:

```
Cluster centroids:
                                  Cluster#
Attribute         Full Data        0          1          2
                     (90)         (30)       (33)       (27)
========================================================
JENSIAT               M            M          F          M
SEN                 23.4         22.3     22.9697    25.1481
MOADEL            15.3222        14.4         16     15.5185
TPEDAR                B            B          B          A
TMADAR                B            B          B          A
ZAMANBANDI            C            C          C          C
ETEMADBENAFS          B            B          A          A
TAMOL                 A            B          A          A
SENEF                 C            B          C          C
MAHAL                 A            A          A          B
HADAF                 D            D          A          D
SHOMARE               A            A          A          A
MAZHAB                B            B          B          A
TASOB                 A            A          B          A
FARHANG               C            C          B          C
SAFAR                 B            B          B          B
KHANEVADE             C            C          C          C
TASALOT               B            M          G          B


Clustered Instances

0        30 ( 33%)
1        33 ( 37%)
2        27 ( 30%)
```

Fig.2. Created clusters by applying clustering and K-means method on data

Qualities of students who were successful in learning English language:
- Gender of majority of them was female.
- Their mean GPA was 16.
- Their parents' education was same and high school diploma.
- They had high self-confidence.
- Their social interaction was good.
- Their age of language learning was 12-18.
- They had learned English language in language schools.
- Their goal for learning language is personal interest.
- They had been bounded to religion as much as possible.
- They were middling prejudiced on their national culture.
- They were middling acquainted with culture of English language countries.

Qualities of students who were mean in learning English Language:
- Gender of majority of them was male.
- Their mean GPA was 14.
- Their parents' education was same and high school diploma.
- They had mean self-confidence.
- Their social interaction was mean.
- Their age of language learning was 6-12.
- They had learned English language in language schools.
- Their goal for learning language is that it is useful in job market and education continuation.
- They had been bounded to religion as much as possible.
- They were highly prejudiced on their national culture.
- They were slightly acquainted with culture of English language countries.

Qualities of students who were unsuccessful in learning English Language:
- Gender of majority of them was male.
- Their mean GPA was 15.
- Their parents' education was same and under high school diploma.
- They had high self-confidence.
- Their social interaction was good.
- Their age of language learning was 12-18.
- They had learned English language in school.
- Their goal for learning language is that it is useful in job market and education continuation.
- They had been bounded to religion completely.
- They were highly prejudiced on their national culture.
- They were slightly acquainted with culture of English language countries.

6.2. Results of Data Processing by Classification and Decision Tree Algorithm Method

In classification method we have groups whose entity has been determined beforehand and we place all objects in their relevant group one by one. In decision tree algorithm, samples are transferred according to their quantities in tree and this will continue until there is no more possibility for making sub-branches. The resulted tree from studied data had

many branches and leaves; after omitting some ineffective fields and omitting nodes which had less than 5 objects in them, the following decision tree was obtained:
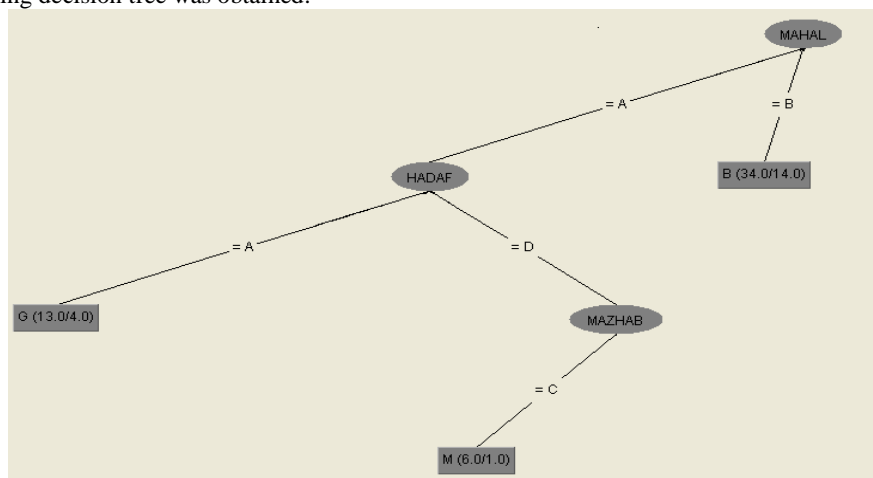
Fig.3. Decision Tree

Results obtained from decision tree:
- Students who studied English language at school were unsuccessful.
- Students who studied English language in language institutes and their aim were personal interest and they have been successful in learning English language.
- Those who studied English language in language institutes and their aim were usefulness of English language in job market and education continuation, had their own religious beliefs and had a mean condition in learning English language.
- Place of learning English language have been a very essential point in learning language for students, in a way that those who studied language in language institutes have been successful and those who studied English language in schools were unsuccessful.

6.3. Conclusion and Future Works

Studying the results of both methods we can conclude that the most significant and effective point in university students' success in learning English language was place of learning language. In a way that students who studied English language in language institutes have been successful and those who studied English language in schools have been unsuccessful in learning. Considering the obtained results we can think about the point that the effect of language institutes on students' success was significant and language learning condition in schools have had some problems and for improving learning in university students we can manage to mobilize schools for teaching language and making students interested in learning English language and using language institutes.

REFERENCES

1. Kahmatian, R. and Atrashi, M., Considering the Role of Gender in Foreign Language Learning Process, Human Sciences 2007. No. 55, P. 147-158.

2. Zareei Tajareh, A., Studying the Role of Age, Gender and Previous Education of English Language Students in Their Success in Language Learning, Esfahan University, 1983.

3. http://www.jamewhehkhedmatkangavar.blogsky.com

4. Sheikhi, M., Factors Affecting the Life Expectancy with Science Data Mining. The fifth Iran Data Mining Conference, Amirkabir University of Technology, Tehran, Iran 2011.

5. Ahmadpour Kasgary, A., Risk Anticipation Model in Choosing Companies Present in Stock Market Using Data Mining, The fifth Iran Data Mining Conference, Amirkabir University of Technology, Tehran, Iran 2011.