

# Identifying Trends Concerning Computer Science Students with Dyslexia: A Data Mining Approach

M Saracee, C. Edwards, A. Peers, A. Redfern, P. Wilkinson  
School of Computing, Science and Engineering  
University of Salford, UK

## ABSTRACT

With dyslexia becoming more widely accepted as a learning disability, many questions come to light about what is being done to help those affected by it. Are those diagnosed given enough support in early education and are they being encouraged to continue with their studies at a higher level? With dyslexia being a learning disability normally specific to the use of language, it is thought that those with the disability often excel at subjects that are less language dependent, one of these subjects being Computer Science. This paper documents the study of data collated from students on a Computer Science related degree at Salford University and University of Manchester (UMIST) with a the goal of identifying any trends, specifically focusing on those concerning students with dyslexia. To identify the trends we, in the end, decided to use Decision Trees and Association Rules and applied them to the dataset derived from the collected questionnaire answers. Statistical techniques were also applied, including graphs, to the same dataset. The statistical results we obtained met our expectations while parts did not which we thought to be interesting. The results from the mined data that we obtained gave us quite a few rules but we picked those that had a reasonably high support and/or confidence and if the rule(s) make sense of which was also a choosing factor even if the support and confidence were not that good.

## Keywords

Medical Data Mining, Dyslexia, Computer Science, Data Mining Techniques, Learning Disabilities, Questionnaire, Decision Trees, Association Rules, Statistics

## 1. INTRODUCTION

Dyslexia affects the lives of many adults and children alike. For those who suffer with this disability many tasks can prove quite difficult even if others find them simple. Most people have heard of the term dyslexia but its true meaning is often misunderstood. The actual word 'dyslexia' comes from the Greek language and is roughly translated as 'poor language.' With dyslexia, the brain's ability to translate signals received from the eyes and/or ears into

An understandable language is impaired. Dyslexia sufferers can have a wide range of symptoms; they are not limited to the simple reading/writing difficulties that are commonly associated with the disability. Other symptoms can include problems with short-term memory, concentration, mathematics and personal organization. If not diagnosed, dyslexia can cause a sufferer to feel less intelligent than their peers when often this is not the case. If left untreated in younger children it can often lead to behavioral difficulties or low self-esteem. Luckily with the increase in awareness of this disability, teachers are becoming more educated towards recognising its symptoms and an early diagnosis is more likely to occur. If treated at an early age the sufferer is more likely to show vast signs of improvement but any dyslexia sufferer can learn to work around their difficulties.

Those who suffer with dyslexia have a different way of learning to others. Although this hinders their understanding and use of language, it may mean that they excel at subjects less language dependant. With the right support in education those with the disability can go on to be very successful with their studies and should have no problems getting into and succeeding in higher education. What we would like to do is study data from students in higher education on a course that a dyslexic is more likely to succeed with. The courses that we have selected to use are those of the Computer Science discipline, of which may give a possible trend between the courses (and other related aspects) and dyslexic students. Hopefully the data collated will give insight as to finding out if enough support is given to dyslexic students in education and if they are given encouragement to enter into higher education. We are also interested in finding out any correlations in the data to see what types of students are dyslexic and how they entered into their degree.

To find patterns in the data collated from students, data mining techniques were used. Data mining can also be called 'knowledge-discovery' and is a process that searches for patterns within a database. The process can be done manually by analysing the data until a pattern becomes apparent or it can be done automatically with the use of programs that analyse the data. This technique will be useful

to this study as it can help find patterns in the data we collect that manually will be laborious and may take a long time.

The rest of this paper will be structured as follows; Section 2 will give an outline of the problem that we are trying to solve, followed by details of other work carried out in this area. Section 3 provides details of the methodology and techniques used in the experiment. Section 4 will be an overview of the work carried out in the experiment with the results and a discussion. Conclusions will be contained in section 5 and finally, section 6 will contain any references used in this paper.

## **2. PROBLEM STATEMENT**

The aim of this study is to apply existing data mining techniques to data obtained from university students on a Computer Science related degree in order to obtain details of any trends specific to dyslexia. We are interested in finding out how the disability affects students and if there are any specific groups of people that are more or less likely to have the disability. The results obtained should provide an insight into whether enough is being done to support those with dyslexia and whether or not dyslexia sufferers are encouraged to enter into higher education.

### **2.1 Related Work**

Much work has been carried out into the area of dyslexia and learning disabilities. Looking specifically into studies around dyslexia in higher education, there have been many interesting findings. The British Psychology Society carried out an interesting investigation into dyslexic students in higher education and whether their disabilities impose a greater impact on their study skills compared to non-dyslexic student. It was found that while non-dyslexic students performed slightly above average; students with dyslexia only attained the average level for their age, which was considerably poorer than expected in relation to their IQ. Their results indicated that there was a great need for support in the development of writing skills, which needed to be flexible and subject specific, rather than in a general form. The study also noted that although dyslexia awareness and support has recently increased, more needs to be done to raise the understanding of the disability.

Richardson and Waydell (2003) also obtained similar results in their study. They found that students with dyslexia were more likely to withdraw from university within their first year and less likely to complete their course

compared to peers without dyslexia. If the dyslexic student went on to complete their degree, they were more likely to gain a lower class than others. They state that although dyslexia can be potentially harmful towards the success of a degree, with the right level of support and resources, there is no reason why a dyslexic student cannot be successful at degree level. Specific to dyslexia in computer science, the paper by Powell et al (2004) documents findings that having dyslexia can actually be beneficial in this area. A dyslexic student can use their different visualisations to tackle programming and may be better adapted to overcome difficulties in learning to program.

There appears to be no documentation of data mining techniques being used to carry out any studies into dyslexia and so this provides an opportunity for interesting findings.

## **3. METHODOLOGY**

This section will detail how the experiment will be carried out. It will cover how the data needs to be collated and which data mining techniques will be used to extract the results.

Firstly, a questionnaire will be sent to all students at UMIST and Salford University on a Computer Science related degree. The questionnaire will be in an electronic form and if the student is willing to take part, they can complete it in their own time. It will ask questions related to dyslexia, their degree and their background. All data collected from the questionnaire will automatically be sent to and stored in an Access database. The constraints on the questions are that they need to none personal.

After the results from the questionnaire have been collated, a dataset needs to be taken. Although the actual size of this dataset will depend upon the number of results received from the questionnaire, the estimated size will be around 50 records. The rest of the data will then be cleaned and any necessary queries will be made.

The final step in the methodology is the actual mining of the data. This will be performed using the data mining tool Envisioner, developed by Neurosoft. The data mining techniques, which will be used, are decision trees and clustering. Statistics will also be studied to see if any other patterns can be seen.

## **4. EXPERIMENTAL EVALUATION**

This section will document the experiment including; what we did, the results that we obtained and exactly what the results show us. In order to collect the relevant data a questionnaire was developed. The website was live for around a month this was considered to be long enough because the amount of responses to the questionnaire were getting less and less as time passed on. The results obtained were based on a set piece of data from the entire data that was collected. From the surveyed information it was possible to produce statistical information, which has subsequently been included in Chart form.

Information provided by the Equality and Diversity Department at Salford University the majority of registered disabled students are of the type dyslexic. Some 410 students are registered [dyslexic] and some 18000 students are registered in the university (estimation) this equates to around 2.25 percent of students having this condition. In addition this department has confirmed that a significant number of students are not diagnosed until they reach university, also the majority of resources this department has, are used on this condition.

### **4.1 Statistical Results**

These results were finalized as graphs; certain aspects of the data were chosen to make comparisons between in particular the data that seems to be related one way of another.

### **4.2 Mining Results**

These results are shown as a Decision Tree and a Report by Envisioner for the Association rules; Dyslexia was chosen for the root of the tree and a basis for the rules as the attribute seems most appropriate and specific attributes where chosen that we thought are the most interesting in correspondence with dyslexia.

### **4.3 Discussion**

#### **Dyslexic students and the different courses**

More dyslexic students take the CSIS Route deducing that this route has less programming and therefore requires less logic and understanding of the programming concept. The dyslexic student relies a great deal on memory and not logic in approaching any task although they are able to see the overall picture they find it difficult to execute the individual stages towards a final solution. Appendix A.

#### **Dyslexia V's Conventional root**

Those students that took a more conventional route in education, i.e. A-Level the Degree compared to those dyslexic students whom took the less conventional route, i.e. Access Course, ND, GNVQ, and HND to the degree. In addition dyslexic students, on un-conventional route, tend to be of a mature type. It can be deduced that this route tends to be more assignment based rather than examination which would suit the dyslexic student, or that this condition has only been diagnosed later in life rather than at the primary or secondary school. However there are cases where dyslexic students have taken the conventional route and those non-dyslexic students whom taken the less conventional.

#### **More Dyslexia through the course**

A progression of the number of dyslexic students with each year of study became apparent; this anomaly could be due to this condition being diagnosed whilst attending university with a subsequent degree of difficulty arising with each year of progression.

#### **Understanding of programming aspects**

Appears that more non-dyslexic students have difficulty in programming logic, syntax and combination as the results appear to show. This actually contradicts chart ones and twos findings. Thus it is fair to think that maybe dyslexic students may understand these aspects more but have difficulty applying the skills.

#### **Dyslexia Vs Ethnic background**

There are fewer dyslexic ethnic students than white students. Reasons here may be due to the reluctance of the ethnic students to make known any such condition or some ethnic students may be from over seas and have a degree of difficulty with the language and may be receiving support for this. In addition this situation may just be that the ethnic student may be harder working due to ability and an ambitious nature.

## **5. CONCLUSION**

In conclusion it gives an insight into dyslexia cases at university and how it might not be not diagnosed until the student progress through the course, see appendix C. This graph shows the progression of each year academic of study.

It would be interesting to look further into obtaining the number cases reported in each year of study, these could be past and present to see if any relationship could be found. To do this we would have to liaison with the equality and diversity department see if a further assistance could be given to these students

particular in diagnosing earlier rather than later so that the overall results of individual students could be improved

There is no real evidence showing a difference of ethnic and white students having this condition although it does point to more white students diagnosed or registered for this condition. This again would be an interesting subsequent study to see if this theory could be substantiated or disproved. With regard to the results produced by the Envisioner the association rules only produced the following that were considered to be of higher enough support and confidence.

In the Envisioner report the first highlighted rule shows, that students who took the conventional route to get to university e.g. 'A' level's to university, were less likely to have dyslexia. The support being 74 percent with a confidence of over 94 percent this also is confirmed by the statistics graph.

The second rule confirms the above with regard to students taking the non-conventional route where more likely to have dyslexia. The third highlighted rule shows that the white students are more likely to be dyslexic although the charts confirm this the results show a support of 54 percent and confidence 22 percent this is not conclusive

Rules four and five highlighted show no difference in whether the white student as understanding of computing language e.g. syntax and logic

Rule six (highlighted) shows the mature student is more likely to be dyslexic this is confirmed with the statistical information that concluded the student taking the non-conventional route is also more likely to have this condition.

## 6. REFERENCES

- [1] Griffiths, Y.M., Hatcher, J and Snowling, M.J. "Cognitive Assessment of Dyslexic Students in Higher Education", *British Journal of Educational Psychology* (2002), 72, 119–133
- [2] Richardson, J.T.E., and Wydell, T.N. "The Representation and Attainment of Students with Dyslexia in UK" *Higher Education. Reading and Writing* (July 2003) 475-503.
- [3] Powell, N. et al. "Dyslexia and Learning Computer Programming", Leeds Metropolitan University 2004.
- [4] The International Dyslexia Association. *Dyslexia Basics. Fact sheet 62* (May 2000)

- [5] *Dyslexia and Related Disorders. Fact Sheet 63*, The International Dyslexia Association. (January 2003)
- [6] Cercone, N. "What Data Mining is". Computer Science Department, University of Waterloo. (2002)
- [7] *Microsoft Access 2003 Inside out* Published by Microsoft press, John L Viescas © 2004
- [8] Bhavani Thuraisingham, *Data Mining: Technologies, Techniques, Tools, and Trends*, 1999.
- [9] Jiawei Han, Micheline Kamber, *Data Mining: Concepts and Techniques*, Academic Press © 2001
- [10] D.R. Cox and E.J. Snell, *Applied Statistics: Principles and Examples*, Chapman and Hall 1981