Using machine learning advances to unravel patterns in subject areas and performances of university students with special educational needs and disabilities (MALSEND): a conceptual approach

Sobnath, Drishty, Isiaq, Olufemi, Rehman, Ikram ORCID logoORCID: https://orcid.org/0000-0003-0115-9024 and Nasralla, Moustafa (2020) Using machine learning advances to unravel patterns in subject areas and performances of university students with special educational needs and disabilities (MALSEND): a conceptual approach. In: Springer ICT Congress (ICICT), 25-26 Feb 2019, London, UK.

http://dx.doi.org/10.1007/978-981-32-9343-4_41

A Machine Learning Conceptual Approach to Detect Patterns in Subject Areas and Performance of University students with Special Educational Needs and Disabilities (MALSEND)

Sobnath.D[1], Isiaq.O[1], Rehman.I[2], Nasralla.M[3]

[1] Research Innovation and Enterprise, Solent University, Southampton, UK

[2] Department of Engineering, Coventry University Group, Coventry, UK

[3] Department of Communications and Networks Engineering, Prince Sultan University, Riyadh, Saudi Arabia

Abstract: Universities and colleges in the UK welcome almost 30,000 disabled students each year. Research shows that the dropout from education in the EU for the disabled is at 31.5%, much higher compared to only 12.3% for non-disabled students. Supporting young students who require special educational needs in pursuing higher education is an ambitious and necessary step that needs to be adopted by tertiary education providers worldwide. We propose, MALSEND, a project aiming to develop a platform based on machine and human intelligence to understand learning disability patterns in Higher Education. The platform will analyse datasets from universities in the previous years and will help to discover any trends in subject areas and performance among autistic students, dyslexic students or students having attention deficit hyperactive disorder (ADHD), among others. Analysing variables such as students' courses, modules, performances and other engagement-indices will give new insights on research questions, career advice and institutional policy making. This paper describes the activities of the development phases of this concept.

Keywords: MALSEND, machine learning, special educational needs, performance, unsupervised learning

1 Introduction

The term "Special Educational Needs and Disabilities" (SEND), refer to students who have learning problems or disabilities that make it harder for them to learn than most of their peers. This may include physical, development disabilities, behavioural, emotional and communication disorders and learning deficiencies (Kryszewska, 2017). Universities and colleges in the UK welcome almost 30,000 disabled students each year (UCAS, 2018). As of today, there are only 9 EU countries, including France and the United Kingdom, who have implemented policy plans to help SEND students in higher education (Limbach-Reich & Powell, 2016). Some of these include free transport to and from universities, special software to aid learning and teaching and other simple assistance to students with specific impairment. However, there is a lack of support and social inclusion for students having a learning disability. There are a number of other concerns such as poor quality, wrong career advice or lack or guidance for students with a learning disability (Disability Rights UK, 2017). Research shows that the dropout from education in the EU for the disabled is at 31.5%, much higher compared to only 12.3% for non-disabled students (Limbach-Reich & Powell, 2016). Supporting young students who require special educational needs in pursuing higher education is an ambitious and necessary step that needs to be adopted by tertiary education providers worldwide. Figure 1 shows the number of people in the UK with a learning disability.

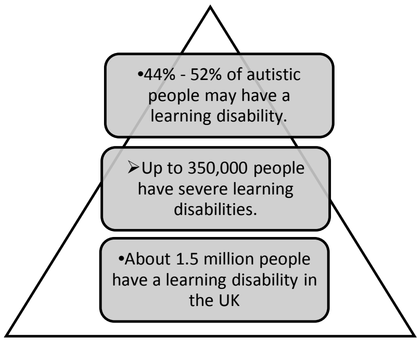

Figure 1: Number of people with learning disabilities in the UK (NHS, 2018)

MALSEND project aims at building an artificial intelligence platform using machine learning algorithms to identify any patterns in the type of learning disability of a student and their chosen subject at university level. The platform will analyse large datasets from UK universities in the previous years and will be trained to discover any trends in subject areas and performance among autistic students, dyslexic students or students having attention deficit hyperactive disorder (ADHD), among others. The system might also help us identify other trends among certain learning disabilities and subject performance based on big datasets which would be impossible for humans to detect otherwise.

1.1 Statement of the Problem

Disability Rights UK states that there are a number of issues they found out during their events, steering groups and participation in various meetings and seminars, among which were the issues of having poor quality or wrong career advice for students with a learning disability, bad choice of subjects that do not match these students' aspirations and lack of guidance for students with specific learning disabilities (Disability Rights UK, 2017). There is a lack of support and social inclusion for students having a learning disability to go into higher education, participate in various domains and contribute to the economy of a country (Kim, Shin, Yu, & Kim, 2016) . This might include having a regular income, having a good quality of life and having access to the overall social and educational system. Pathways to tertiary education for SEND students depend on the type of learning disability, financial resources and self-motivation of students after leaving secondary schools. To reduce dropout and encourage those with learning disabilities, research must be carried out to see what are the best options for these students in terms of subject areas, performance and their potential contribution to the society. However, there are no models that can detect any trends or patterns in the learning disability of a student, their performance and success in specific areas. If such a model can be implemented, then the probability of having more students to enter tertiary education might rise and hopefully reduce the dropout rate since the latter will know that similar students suffering from the same learning disability managed to complete their degree. Having such a model can also help students to have a voice and help organisations in policy making around SEND students.

1.3 Purpose of Inquiry and Inquiry Questions

This conceptual paper is proposed to identify and recognise factors that affect the learning and performance of higher education students suffering from a learning disability through an analysis of students' records from previous years. Therefore, the research questions for this study focuses on the following; 1) Can a machine learning platform be used to identify any repeated patterns of data among a learning disability, subject area and performance of students at tertiary education level? and 2) Can the platform be trained with more datasets to then predict the subject area a SEND student is most likely to succeed based on his learning disability?

The next section of this paper discusses the related work.

2 Related Work

2.1. Machine Learning

Machine Learning (ML) is an application of artificial intelligence (AI) whereby programs uses statistical models to give computer systems the ability to learn without being fully programmed (Jordan & Mitchell, 2015). The adoption of machine-learning methods can be found throughout different areas to improve marketing, decision-making in health care, manufacturing, education and also applied for financial analysis. For example, machine learning is being used to predict mortality rate associated with certain type of diseases, predict effectiveness of surgical procedures, to help physicians make better decision and to discover relationship among clinical and diagnosis data. Unsupervised algorithms attempt to overcome limitations of supervised learning algorithms by automatically identifying patterns and dependencies in the data. Unsupervised learning can be beneficial for this study as it allows the algorithms to look back for patterns that have not been previously considered. Useful new information can therefore be extracted from the observed data to building predictors. In statistical unsupervised learning pattern recognition, the data can be identified by finding clusters, for example by using K-means algorithms or Adaptive Resonance Theory (ART) adapted algorithms, dimensionality reduction or Principle Component Analysis (PCA) to reduce the number of random variables under consideration by defining a set of principle variables (Wagstaff, Cardie, Rogers, & Schroedl, 2008). It is important to understand the relation between identified clusters and competitive learning algorithms can provide efficient solutions to problems.

2.2 Prediction of student's performance

Previous work has been done to predict the performance of students as shown by a few studies (Cortez & Silva, 2008)(Thiede et al., 2015)(Chamorro-Premuzic & Furnham, 2003). Different machine learning techniques, such as matrix factorization (Thai-Nghe, Horváth, & Schmidt-Thieme, 2011) or collaborative filtering (Toscher & Jahrer, 2010) have been used to predict students' grades. The right research questions are important to understand the existing studies of predicting SEND students' area of expertise. Current studies make use of Cumulative Grade Point Average (CGPA), assignment mark, quizzes, lab work, class test and attendance to predict performance of students in general (Mohamed Shahiri, Husain, & Abdul Rashid, 2015). Other researchers considered gender, age, family background, disability, extra-curricular activities, social interaction and psychometric factors (Mohamed Shahiri et al., 2015) to see how these affect the student's performance. The proposed study, however, is looking at identifying relationships between identified factors and SEND students' areas of expertise by analysing at least 15,000 student records. To the best of the authors' knowledge, there are no studies that tried to identify how a specific learning disability can be associated with particular subject areas and this is an identified gap in the literature.

3 MALSEND Platform

A conglomerate of approaches will be adopted at different stages of the research work. Multiple anonymised datasets of student records from universities will have to be examined to determine existing patterns. However, for the purpose of this conceptualisation, we are only considering students' data from 2 UK universities. Once the platform has been implemented and initial results have been obtained, the platform will be reinforced and trained with datasets from other universities. The following sections describe the activities of the development phases of this concept.

A. Ethical Approval

This work is being carried out under strict ethical standards, for example in relation to students' privacy, confidentiality and university's consent. Ethical approval has therefore been obtained for this research project from the participating universities' Ethics Committee in November 2018. The data collected will be completely anonymised to prevent the identification of any student and to abide by the General Data Protection Regulations (GDPR) EU regulations. Another ethics application will be made in the second stage of the project when data from other universities will be required to reinforce and test the platform.

B. Datasets

At least 15,000 anonymised student records over the last 8 years from 2 UK universities will be analysed for the first pilot study. Anonymised data for students who have been clinically diagnosed with one of the published learning disabilities (dyslexia, dyspraxia, ADHD, Asperger's syndrome, other autistic spectrum disorder), as shown in Table 1, is being collected in a spreadsheet. In the UK, Higher Education institutions uses the follow standard codes to classify disabilities, a coding frame introduced by the HESA and the Disability Rights Commission (DRC) (HESA, 2016)

Table 1: Type of learning disability recorded by HE institutions in UK

Code	Label
0	No known disability
8	Two or more impairments and/or disabling medical conditions
51	A specific learning difficulty such as dyslexia, dyspraxia or ADHD
53	A social/communication impairment such as Asperger's syndrome/other autistic spectrum disorder
54	A long-standing illness or health condition such as cancer, HIV, diabetes, chronic heart disease, or epilepsy
55	A mental health condition, such as depression, schizophrenia or anxiety disorder
56	A physical impairment or mobility issues, such as difficulty using arms or using a wheelchair or crutches
57	Deaf or a serious hearing impairment
58	Blind or a serious visual impairment uncorrected by glasses
96	A disability, impairment or medical condition that is not listed above

The data collected such as age range (e.g. 19-21 years old, 22-25 years old), sex, status (full time, part time, distance learners), type of learning disability (autistic, ADHD, dyslexic, dysgraphia), entry type (foundation/A-level/diploma), A-Level of students (UCAS points and subjects), module grades, no of sittings, no of credits, module type, course code and description, module code and description, whether the student is an undergraduate or postgraduate student, results in 1st, 2nd, 3rd year or postgraduate results, alumni information (career path, job position after graduation) and any other parameters, will be filtered to be analysed by different algorithms as explained in the next section.

C. Analysis

I. Dimensionality reduction

Using Scikit-Learn (software machine library for Python) (Géron, 2017), the data will first be fed to a dimensionality reduction algorithm to reduce the number of variables and simplify the data without losing much information. There are group of algorithms that can be used to remove unneeded data, outliers, and other non-useful data. Dimensionality reduction will be made to free storage space on our server and improve the performance of our machine learning system. It will also help the researchers to visualize the data (Hurwitz & Kirsch, 2018). An anomaly detection algorithm can also automatically remove outliers from the datasets.

II. K-Means Clustering

Clustering (k-means) and visualisations algorithms can then be applied to the dataset to identify clusters and unsuspected patterns. Finally, another method of unsupervised learning, known as association rule learning algorithm will be used to discover interesting relations among other attributes.

The findings will be evaluated with the second dataset in the next stage of the project. The prototype can also be further developed with new data to predict subject areas of SEND students in the future. Figure 3 shows the components of the proposed MALSEND platform.

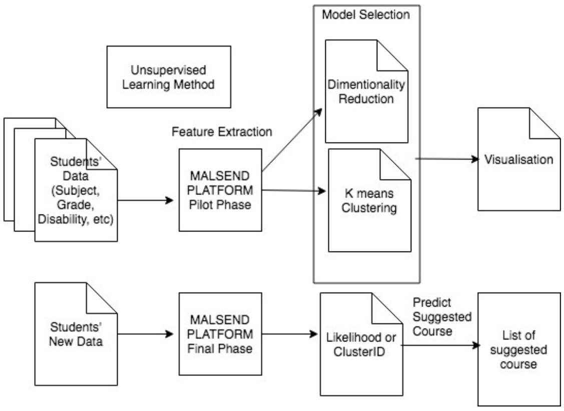

Figure 2: Components of the proposed MALSEND platform

Key algorithms help in model creation to determine patterns, correlations and clusters from the data. The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data.

Limitations

Thre are a few identified limitations for this pilot study which can affect the results. For example, since the sample data is taken from only 2 universities in the first phase, the results does not represent the demographics of the population. Moreover, the fact that some universities provide more specific courses than other universities, the results may show certain patterns and high correlations among a specific learning disability and a course. Hence, there is a need to carry out the second part of the project to test any hypothesis and conclude any findings. More importantly, there are other factors such as social, economic factors or family background that can also affect the results of this study.

Discussion and Conclusion

We expect the machine learning platform to generate new knowledge by identifying patterns that could help solve some of the social challenges such as high dropout rates from education (31.5% compared to 12.3% for non-disabled students), low employability (only 6% of adults with a learning disability in England are in paid work) or depression among SEND students in the future. The findings will create new research questions and help bring other universities together through collective intelligence to find similar patterns regarding other health conditions (visual or hearing impairment, epilepsy). The results in wider applicability could also be used to support career advisers in schools, colleges and communities by providing course suggestions tools. Finally, the findings could assist the government and other institutions for the development of policies, curriculum and practices to help those students to find a direction in regards to their career.

As to the best of our knowledge, no such system has yet been implemented to help SEND students in the UK find out what are the subject areas chosen by other past students who had a similar learning disability as them, their success/failure in particular areas and their future career paths. The findings of this study might open new opportunities and act as a guide to those having a learning disability and entering higher education. It might also help in the investigation of the performance of students with a specific learning disability based on identified patterns.

Future AI and machine learning prediction models can provide an extra set of eyes and ears for SEND students, therapists, teachers as well as parents. Further analysis of this data can lead to pinpointing social success factors and assessing a student's strengths and weaknesses. Moreover, educators would be better able to understand the students' learning and emotional development, slowly introducing them to more complex and varied social environments over time.

Acknowledgements

This work is supported by the Research Innovation and Enterprise department at Solent University.

References

1. Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality*. https://doi.org/10.1016/S0092-6566(02)00578-0
2. Cortez, P., & Silva, A. (2008). Using Data Mining To Predict Secondary School Student Performance. *In the Proceedings of 5 Th Annual Future Business Technology Conference*. https://doi.org/10.13140/RG.2.1.1465.8328
3. Disability Rights UK. (2017). *Careers Guidance And Advice For Disabled Young*

People. Retrieved from https://www.disabilityrightsuk.org/sites/default/files/pdf/Careers Guidance And Advice For Disabled Young People.pdf

4. Géron, A. (2017). Géron - 2017 - Hands-On Machine Learning with Scikit-Learn and TensorFlow.pdf. In *Hands-on Machine Learning with Scikit-Learn and TensorFlow*. https://doi.org/10.3389/fninf.2014.00014
5. HESA. (2016). Fields required from institutions in All fields Disability. Retrieved January 10, 2019, from https://www.hesa.ac.uk/collection/c16051/a/disable
6. Hurwitz, J., & Kirsch, D. (2018). *Machine Learnning For Dummies, IBM Limited Edition Published*. (C. Burchfield, Ed.). John Wiley & Sons, Inc.
7. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255 LP-260.
8. Kim, K. M., Shin, Y. R., Yu, D. C., & Kim, D. K. (2016). The Meaning of Social Inclusion for People with Disabilities in South Korea, *64*(May), 19–32. https://doi.org/10.1080/1034912X.2016.1165802
9. Kryszewska, H. (2017). Teaching Students with Special Needs in Inclusive ClassroomsSpecial Educational Needs. *ELT Journal*, *71*(4), 525–528. https://doi.org/10.1093/elt/ccx042
10. Limbach-Reich, A., & Powell, J. (2016). *Young adults with special educational needs (SEN)*. Retrieved from http://www.deutschlandfunk.de/bachelor-und-master-wir-
11. Mohamed Shahiri, A., Husain, W., & Abdul Rashid, aini. (2015). ScienceDirect The Third Information Systems International Conference A Review on Predicting Student's Performance using Data Mining Techniques. *Procedia Computer Science*, *72*, 414–422. https://doi.org/10.1016/j.procs.2015.12.157
12. NHS. (2018). Learning disabilities - NHS. Retrieved August 23, 2018, from https://www.nhs.uk/conditions/learning-disabilities/
13. Thai-Nghe, N., Horváth, T., & Schmidt-Thieme, L. (2011). Factorization models for forecasting student performance. In *Proceedings of the 4th International Conference on Educational Data Mining*. https://doi.org/10.1016/0092-8674(94)90138-4
14. Thiede, K. W., Brendefur, J. L., Osguthorpe, R. D., Carney, M. B., Bremner, A., Strother, S., … Jesse, D. (2015). Can teachers accurately predict student performance? *Teaching and Teacher Education*. https://doi.org/10.1016/j.tate.2015.01.012
15. Toscher, a, & Jahrer, M. (2010). EDM-59: Collaborative filtering applied to educational data mining. *Austria - KDD Cup*.
16. UCAS. (2018). Disabled Students | Advice And Financial Support | UCAS. Retrieved September 19, 2018, from https://www.ucas.com/undergraduate/applying-university/individual-needs/disabled-students
17. Wagstaff, K., Cardie, C., Rogers, S., & Schroedl, S. (2008). Constrained K-means Clustering with Background Knowledge. *International Conference on Machine Learning*. https://doi.org/10.1109/TPAMI.2002.1017616