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

Drishty Sobnath, Olufemi Isiaq, Ikram Rehman, Moustafa Nasralla

Abstract: Universities and colleges in the UK welcome about 30,000 students with special needs each year. Research shows that the dropout rate for disabled students is much higher at 31.5% when compared with about 12.3% for non-disabled students in the EU. Supporting young students with special educational needs while pursuing higher education is an ambitious and important role, which needs to be adopted by tertiary education providers worldwide. We propose, MALSEND, a conceptual platform based on Human Machine Intelligence (HMI), a collective intelligence of human and machine to understand patterns of learning of disabled students in Higher Education. This platform aims to accommodate and analyse datasets features of universities activities to discover trends in performances with regards to subject areas for autistic students, dyslexic students and students having attention deficit hyperactive disorder (ADHD), among others. Analysis of variables, such as students' performances in modules, courses and other engagement-indices will give new insights into research questions, career advice and institutional policymaking. This paper describes the developmental activities of the MALSEND concept in phases.

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

1 Introduction

The term "Special Educational Needs and Disabilities" (SEND), refers 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, only 9 countries of the European Union, including France and the United Kingdom, have implemented policy plans to help SEND students in higher education (Limbach-Reich & Powell, 2016). Some of these plans 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 (Kim, Shin, Yu, & Kim, 2016). 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, when 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 all tertiary education providers worldwide. Figure *1* shows the number of people in the UK with a learning disability.

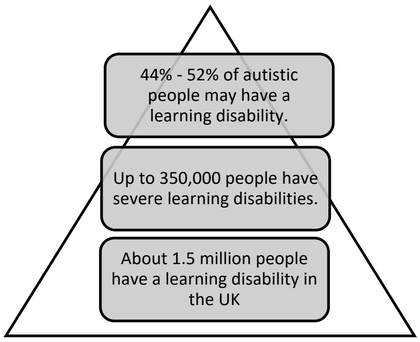

Figure 1: Number of people with learning disabilities in the UK (NHS, 2018)

MALSEND project aims at developing an intelligence platform using learning algorithms to identify learning patterns of disabled students and their corresponding chosen subject at university level. The platform intends to explore large datasets about students-universities activities to discover trends in subject areas and performance among autistic students, dyslexic students or students having attention deficit hyperactive disorder (ADHD) among others.

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. There is also a bad choice of subjects that do not match these students' aspirations and a lack of guidance for these students (Disability Rights UK, 2017). There is a lack of support and social inclusion for students with learning disabilities in going into higher education, participating in various domains and contributing to the economy of a country (Kim, Shin, Yu, & Kim, 2016). Subsequently, this results into problems such as having a regular income, good quality of life and 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 the rate of dropout and encourage students with learning disabilities, research must be carried out to identify effective and suitable options for these students in terms of subject areas, performance and their potential contribution to the society. However, to the authors' knowledge, there is no platform, framework or model available to detect any trends or patterns in the learning disability of a student, their corresponding performances and success in specific subject areas. Implementing such models can improve the possibility of making tertiary institution more conducive and hopefully reduces the dropout rate. Also, such models can help special needs students have a voice in HE and help organisations in policy making relating to SEND students.

1.2 Purpose of Inquiry and Inquiry Questions

This paper is a conceptual proposition for the discovery and identification of factors that affect the learning and performance of higher education students suffering from a learning disability. Using a collective intelligence approach, an extensive analysis of students' records will be utilised to determine relative trends. Therefore, the research questions for this study focus on the following; 1) Will a machine learning platform suitably identify patterns of learning disability, subject area and performance of students at tertiary education level? and 2) Can a Machine learning platform accurately predict subject area performance for SEND students based on their learning disability? The next section of this paper discusses the related work.

2 Related Work

Machine Learning (ML) is an aspect of artificial intelligence (AI) whereby programs uses statistical models to give computer systems the ability to learn without being fully programmed (Jordan & Mitchell, 2015). Nowadays, the application of machine-learning techniques is rampant throughout different areas including marketing improvement, 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. Therefore, this work can benefit from unsupervised learning allowing algorithms to look back for patterns that have not been previously considered. Consequently, new knowledge can 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) algorithms. Dimensionality reduction or Principle Component Analysis (PCA), is another method used 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 so that competitive learning algorithms can be applied to provide efficient solutions to problems.

2.1 Prediction of student's performance

Previously, work has been done in the area of predicting 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. However, the proposed work is looking at finding relationships between identified factors and SEND students' areas of expertise by analysing at least 15,000 student records since this was an identified gap in the literature.

3 MALSEND Platform

A composition of approaches with multifaceted techniques is adopted at different stages of the research work. At the initial stage, multiple anonymised datasets of student records from universities is 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 use the following 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 (HESA, 2016)

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

Data collected consists of 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). Also included are 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 academic level (undergraduate or postgraduate). In addition, grades of 1st, 2nd, 3rd year or postgraduate results, alumni information (career path, job position after graduation) and other related parameters are considered to be analysed using suitable algorithms as explained in the next section.

C. Analysis

I. Dimensionality reduction

Following the Scikit-Learn (software machine library for Python) (Géron, 2017), dimensionality reduction algorithm will be utilised to reduce the number of meaningful variables to simplify the data without losing much information. In addition, other groups of algorithms can be adopted to remove unneeded data, outliers, and other non-useful data. Dimensionality reduction will be performed 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.

III. Components of MALSEND Platform
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 2 shows the components of the proposed MALSEND platform.

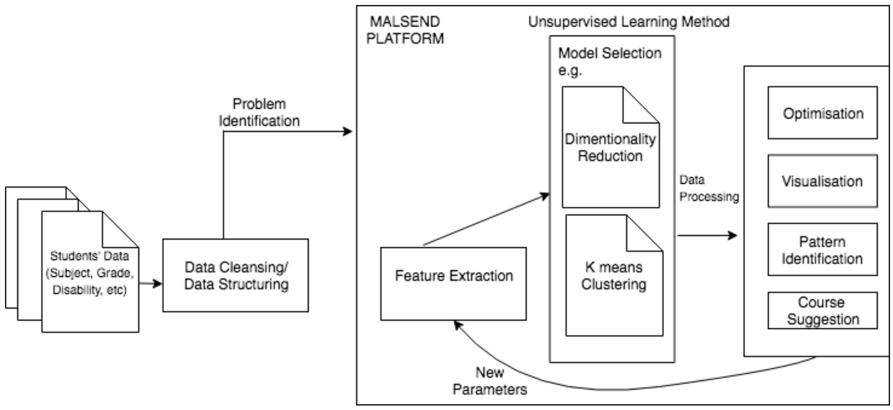

Figure 2: Components of the proposed MALSEND platform

Key algorithms help in model creation to determine patterns, correlations and clusters from the data. The objective for unsupervised learning is to model the fundamental organisation or scattering in the data in order to learn more about it.

4 Limitations

There 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 do 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.

5 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. The findings of this study might open new opportunities and act as a guide to those having a learning disability and who are planning to pursue higher education studies.

As to the best of our knowledge, no such system has yet been implemented to help SEND students in the UK find out which are the subject areas they will most likely to be successful in, based on past students who had similar learning disabilities. Future AI and machine learning prediction models can provide an extra set of eyes and ears for SEND students, therapists, teachers as well as for parents. Further analysis of this data can lead to pinpointing social success factors and assessing a student's strengths and weaknesses. Finally, the findings could assist the government and other institutions for the development of policies, curriculum and educational practices. 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.

6 Acknowledgements

Dr. Drishty Sobnath would like to acknowledge the Research, Innovation and Enterprise department of Solent University for supporting this work. Dr. Ikram Ur Rehman would like to thank the CU Coventry for its support. Dr. Moustafa Nasralla would like to acknowledge the management of Prince Sultan University (PSU) for the valued support and research environmental provision which have led to completing this work.

7 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