A NEW FRAMEWORK IN IMPROVING PREDICTION OF CLASS IMBALANCE FOR STUDENT PERFORMANCE IN OMAN EDUCATIONAL DATASET USING CLUSTERING BASED SAMPLING TECHNIQUES

SULTAN JUMA SULTAN ALALAWI
902668

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2021

Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI

(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

SULTAN JUMA SULTAN AL-ALAWI (902668)

calon untuk Ijazah **PhD**
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"A NEW FRAMEWORK IN IMPROVING PREDICTION OF CLASS IMBALANCE FOR STUDENT PERFORMANCE IN OMAN EDUCATIONAL DATASET USING CLUSTERING BASED SAMPLING TECHNIQUES"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : ***07 Disember 2021.***
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on: ***07 December 2021.***

Pengerusi Viva: *(Chairman for VIVA)*	**Assoc. Prof. Ts Dr Jafri Hj Zulkefli Hew**	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	**Assoc. Prof. Ts. Dr. Mustafa Man**	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	**Dr. Ch'ng Chee Keong**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Ts. Dr. Izwan Nizal Mohd Shaharanee**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Ts. Dr. Jastini Binti Mohd Jamil**	Tandatangan *(Signature)*

Tarikh:
(Date) ***07 December 2021***

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Portal Pendidikan Oman (OEP) menunjukkan bahawa ketidakseimbangan set data biasanya berlaku dalam menilai prestasi pelajar. Sebilangan besar pelajar adalah berprestasi baik, sementara itu hanya sebilangan kecil tidak menunjukkan prestasi yang baik. Teknik klasifikasi untuk set data yang tidak seimbang boleh menyebabkan ketepatan ramalan yang agak mengelirukan. Ketepatan ramalan keseluruhan biasanya didorong oleh kelas majoriti dengan mengorbankan prestasi yang tidak baik pada kelas minoriti. Objektif utama kajian ini adalah untuk meramal prestasi pelajar yang terdiri daripada pengagihan kelas yang tidak seimbang, dengan memanfaatkan pelbagai teknik persampelan bersama dan beberapa model pengkelasan perlombongan data. Tiga teknik persampelan utama iaitu persampelan minoriti sintetik (SMOTE), persampelan bawah rawak (RUS) dan persampelan berasaskan pengelompokan dibandingkan untuk meningkatkan ketepatan ramalan di kelas minoriti sambil mengekalkan prestasi klasifikasi keseluruhan yang baik. Lima model pengkelasan perlombongan data yang berbeza - J48, Random Forest, K-Nearest Neighbour, Naïve Bayes dan Logistic Regression digunakan untuk meramal prestasi pelajar. Pengesahan silang 10 kali ganda digunakan untuk mengurangkan bias persampelan. Empat matriks digunakan dalam menilai prestasi pengelasan: ketepatan, False Positive (FP), pekali korelasi Matthews (MCC), dan Karakteristik Operasi Penerima (ROC). Set data OEP antara 2018 dan 2019 diambil untuk menilai keberkesanan teknik persampelan dan juga kaedah klasifikasi. Hasil menunjukkan bahawa *K-Nearest Neighbour* yang digabungkan dengan teknik persampelan berasaskan pengelompokan menghasilkan prestasi klasifikasi terbaik dengan nilai MCC 98.4% pada pengesahan silang 10 kali ganda. Teknik persampelan berasaskan pengelompokan meningkatkan prestasi ramalan keseluruhan untuk kelas minoriti. Di samping itu, pemboleh ubah terpenting untuk meramalkan prestasi pelajar dengan tepat dikenal pasti dengan menggunakan model *Random Forest*. OEP mengandungi sejumlah data yang besar dan analisis berdasarkan data yang besar dan kompleks ini dapat memberi manfaat kepada pemegang taruh dalam OEP dalam meningkatkan prestasi pelajar dan mengenal pasti pelajar yang memerlukan perhatian tambahan.

Kata Kunci: Perlombongan data, Set data tidak seimbang, Teknik Persampelan, Pengelasan, Pembolehubah terpenting

Abstract

According to the Oman Education Portal (OEP), data set imbalances are common in student performance. Most of the students are performing well, while only small cases of students are underperformed. Classification techniques for the imbalanced dataset can yield deceivingly high prediction accuracy. The majority class usually drives the overall predictive accuracy at the expense of having abysmal performance on the minority class. The main objective of this study was to predict students' performance which consisted of imbalanced class distribution, by exploiting different sampling techniques and several data mining classifier models. Three main sampling techniques – synthetic minority over-sampling technique (SMOTE), random under-sampling (RUS), and clustering-based sampling were compared to improve the predictive accuracy in the minority class while maintaining satisfactory overall classification performance. Five different data-mining classifiers - J48, Random Forest, K-Nearest Neighbour, Naïve Bayes, and Logistic Regression were used to predict the student performance. 10-fold cross-validation was utilized to minimize the sampling bias. The classifiers' performance was evaluated using four metrics: accuracy, False Positive (FP), Matthews correlation coefficient (MCC), and Receiver Operating Characteristic (ROC). The OEP datasets between 2018 and 2019 were extracted to assess the efficacy of both sampling techniques and classification methods. The results indicated that the K-Nearest Neighbors combined with the clustering-based sampling technique produced the best classification performance with an MCC value of 98.4% on the 10-fold cross-validation. The clustering-based sampling techniques improved the overall prediction performance for the minority class. In addition, the most important variables to accurately predict student performance were identified by utilizing the Random Forest model. OEP contains a large amount of data and analyses based on this large and complex data can be useful for OEP stakeholders in improving student performance and identifying students who require additional attention.

Keywords: Data mining, Imbalanced dataset, Sampling technique, Classification, Importance variable

Acknowledgement

First, I would like to thank my wife and my five children Juma, Bissan, Tamim, Gassan and Qais for consistently giving me courage and inspiration. To my parents, who have been instrumental in guiding my life and encouraging me to succeed, I express my deepest thanks. They are the reasons that I am who I am today.

I would like to express my appreciation and gratitude to everyone who has contributed in completing this thesis. It was my pleasure to study under Ts. Dr. Izwan Nizal Bin Mohd Shaharanee supervision. It is not enough to thank you very much to his for his guidance to help me to achieve my goal. Without his valuable support, my thesis would not have been possible. I would like to express my thanks to my co-supervisor Ts. Dr. Jastini Mohd Jamil for her comments which help to improve my work.

I am very grateful to the examiners. They were very kind during the viva and during the period of the correction. Additionally their comments have helped to improve this work.

I had a very enjoyable study at Universiti Utara Malaysia (UUM). Not only, does it has a beautiful natural environment but the university also has helpful staff.

Finally, I would like to thank all of my friends for their encouragement during my study.

Table of Contents

List of Tables

List of Figures

CHAPTER ONE

INTRODUCTION

1.1 Introduction

The Sultanate of Oman's educational system had suffered from the lack of qualified teachers and the destruction of learning facilities such as laboratories and libraries for many years. Since 1971, the education sector in the Sultanate of Oman has undergone many reforms (Education, 2011) in improving the teaching and learning environments, facilities, quality and governance. Referring to this reform, the government of Oman, through the Ministry of Education, has been trying to enhance the quality of education in all schools. For example, education and training in 2017 was given a $1.585bn dollar grant to enhance Oman's educational management in implementing its educational sector plan (Oman, 2017). However, the educative situation in the Sultanate of Oman was still not satisfactory. It was in the 43rd position out of 49 countries that joined the Trends in International Mathematics and Science Study (TIMSS) assessments in the 4th and 8th grades in 2015. The current result that was released by the TIMSS report showed that the Sultanate of Oman's score was 425 points (Mullis, 2015). This result revealed that an average performing Omani student (425) scored below the low international benchmark. In general, the report from TIMSS (Mullis, 2015) showed that students in secondary schools in the Sultanate of Oman suffered from low academic performance.

One of the most important markers of the quality of educational improvement is student performance (Alhajraf & Alasfour, 2014; Aulck & Blumenstock, 2016). Increasing the level of student performance is a long- term aim of any government around the world. Underperforming students give negative effects to parents, decision-

References:

Aguiar, Dame, & Ambrose. (2014). Engagement vs Performance : Using Electronic Portfolios to Predict First Semester Engineering Student Retention * Categories and Subject Descriptors The College of Engineering. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (pp. 103-112). ACM.*, 103–112.

Alashwal, H., El Halaby, M., Crouse, J. J., Abdalla, A., & Moustafa, A. A. (2019). The application of unsupervised clustering methods to Alzheimer's disease. *Frontiers in Computational Neuroscience*, *13*(May), 1–9. http://doi.org/10.3389/fncom.2019.00031

Al-balushi, S. M., Ambusaidi, A., & Al-harthy, I. (2015). Students ' performance in TIMSS test in Oman : Effect of cognitive and metacognitive variables and effectiveness of an inquiry-based mobile e- formative assessment package المتغ ا ريثأتو ... Students ' performance in TIMSS test in Oman : Effect of cognitive. *Researchgate.net*, (April 2016). http://doi.org/10.13140/RG.2.1.2870.3124

Alehegn, M., Joshi, R. R., & Mulay, P. (2019). Diabetes analysis and prediction using random forest, KNN, Naïve Bayes, and J48: An ensemble approach. *International Journal of Scientific and Technology Research*, *8*(9), 1346–1354.

Al-farsi, F. S. S., Sima, N., & Shariff, M. (2014). Effects of Educational Indicators on Students ' Performance in the Sultanate of Oman. *International Journal of Technical Research and Applications*, *3*(3), 41–44.

Alhajraf, N. M., & Alasfour, A. M. (2014). The Impact of Demographic and Academic Characteristics on Academic Performance. *International Business Research, 7(4), 92.*, *7*(4), 92–100. http://doi.org/10.5539/ibr.v7n2p92

Alhassan, A., Zafar, B., & Mueen, A. (2020). Predict students' academic performance based on their assessment grades and online activity data. *International Journal of Advanced Computer Science and Applications*, *11*(4). http://doi.org/10.14569/IJACSA.2020.0110425

Ali, A., Shamsuddin, S. M., & Ralescu, A. L. (2015). Classification with class imbalance problem : *Int. J. Advance Soft Compu. Appl, 7(3).*, *7*(3).

Ali, H., Salleh, M. N. M., Saedudin, R., Hussain, K., & Mushtaq, M. F. (2019). Imbalance class problems in data mining: A review. *Indonesian Journal of Electrical Engineering and Computer Science*, *14*(3), 1552–1563. http://doi.org/10.11591/ijeecs.v14.i3.pp1552-1563

Al-Kiyumi, A., & Hammad, W. (2020). Preparing Instructional Supervisors for Educational Change: Empirical Evidence From the Sultanate of Oman. *SAGE Open*, *10*(2). http://doi.org/10.1177/2158244020935905

Al-Lamki, S. M. (2011). The Development of Private Higher Education in the Sultanate of Oman: Perception and Analysis. *International Journal of Private Education*, (1).

Al-maamari, S. (2014). Education for Developing a Global Omani Citizen : Current Practices and Challenges, *2*(3), 108–117. http://doi.org/10.11114/jets.v2i3.399

Almustafa, K. M. (2020). Prediction of heart disease and classifiers' sensitivity analysis. *BMC Bioinformatics*, *21*(1), 1–18. http://doi.org/10.1186/s12859-020-03626-y

Alrajhi, M. N., Alkharusi, H. A., & Aldhafri, S. S. (2016). Learning Processes and Academic Achievement among Omani School Students. *Canadian Center of Science and Education*, *8*(4), 62–71. http://doi.org/10.5539/res.v8n4p62

Alshoaibi, H. (2015). volution of the omani higher education system and economic challenges 1970-2014. *Southern Illinois University Carbondale*, (Summer).

Altujjar, Y., Altamimi, W., Al-turaiki, I., & Al-razgan, M. (2016). Predicting Critical Courses Affecting Students Performance : A Case Study. *Procedia - Procedia Computer Science*, *82*(March), 65–71. http://doi.org/10.1016/j.procs.2016.04.010

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(1). http://doi.org/10.1186/s41239-020-0177-7

Amin, A., Anwar, S., Adnan, A., Nawaz, M., Howard, N., Qadir, J., … Member, S. (2016). Comparing Oversampling Techniques to Handle the Class Imbalance Problem : A Customer Churn Prediction Case Study. *IEEE Access*, *4*(Ml).

Andrew EstabrooksDuke, Duke Taeho Jo, N. J. (2004). A Multiple Resampling Method for Learning from Imbalanced Data Sets. *Computational Intelligence*, *20*(1), 18–37.

Andrić, K., Kalpić, D., & Bohaček, Z. (2019). An insight into the effects of class imbalance and sampling on classification accuracy in credit risk assessment. *Computer Science and Information Systems*, *16*(1), 155–178. http://doi.org/10.2298/CSIS180110037A

Anuradha, C., & Velmurugan, T. (2015). A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance. *Indian Journal of Science and Technology*, *8*(July), 1–12. http://doi.org/10.17485/ijst/2015/v8i

Arum, R. (1996). Do private schools force public schools to compete? *American Sociological Review*, 29–46.

Ashraf, S., Saleem, S., Ahmed, T., Aslam, Z., & Muhammad, D. (2020). Conversion

of adverse data corpus to shrewd output using sampling metrics. *Visual Computing for Industry, Biomedicine, and Art*, *3*(1). http://doi.org/10.1186/s42492-020-00055-9

Awan, S. E., Bennamoun, M., Sohel, F., Sanfilippo, F. M., & Dwivedi, G. (2020). Imputation of Missing Data with Class Imbalance using Conditional Generative Adversarial Networks. Retrieved from http://arxiv.org/abs/2012.00220

Bahety, A. (2014). Extension and Evaluation of ID3 – Decision Tree Algorithm, 1–8.

Baker, R. S. J., & Carvalho, A. M. J. A. De. (2008). Labeling Student Behavior Faster and More Precisely with Text Replays. *In Educational Data Mining 2008.*

Barros, R. C., Basgalupp, M. P., Freitas, A. A., & De Carvalho, A. C. P. L. F. (2014). Evolutionary design of decision-tree algorithms tailored to microarray gene expression data sets. *IEEE Transactions on Evolutionary Computation*, *18*(6), 873–892. http://doi.org/10.1109/TEVC.2013.2291813

Barros, T. M., Neto, P. A. S., Silva, I., & Guedes, L. A. (2019). Predictive models for imbalanced data: A school dropout perspective. *Education Sciences*, *9*(4). http://doi.org/10.3390/educsci9040275

Bayer, J., Byd, H., & Jan, G. (2012). Predicting drop-out from social behaviour of students. *5th International Conference on Educational Data Mining*, (Dm), 103–109.

Bekele, R., & Menzel, W. (2005). A Bayesian Approach to Predict Performance of a Student (BAPPS): A Case with Ethiopian Students. *Algorithms, 22(23), 24.*

Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation Measures for Models Assessment over Imbalanced Data Sets. *Journal of Information Engineering and Applications*, *3*(10), 27–39.

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. *International Conference on Pattern Recognition*, 3125–3128. http://doi.org/10.1109/ICPR.2010.764

Brzezinski, D., Minku, L. L., Pewinski, T., Stefanowski, J., & Szumaczuk, A. (2021). The impact of data difficulty factors on classification of imbalanced and concept drifting data streams. *Knowledge and Information Systems.* http://doi.org/10.1007/s10115-021-01560-w

Bunkar, K. (2012). Data Mining : Prediction for Performance Improvement of Graduate Students using Classification. *In 2012 Ninth International Conference on Wireless and Optical Communications Networks (WOCN) (pp. 1-5). IEEE.*, 3–7.

Cardona, T. A., & Cudney, E. A. (2019). Predicting student retention using support vector machines. *Procedia Manufacturing*, *39*(2019), 1827–1833.

http://doi.org/10.1016/j.promfg.2020.01.256

Carlos Márquez-Vera, Cristóbal Romero Morales, and S. V. S. (2013). Predicting School Failure and Dropout by Using Data Mining Techniques. *IEEE JOURNAL OF LATIN-AMERICAN LEARNING TECHNOLOGIES, VOL. 8, NO. 1, FEBRUARY 2013*, *8*(1), 7–14.

Chawla, N. V, Japkowicz, N., & Drive, P. (2004). Special Issue on Learning from Imbalanced Data Sets. *ACM Sigkdd Explorations Newsletter*, *6*(1), 2000–2004.

Chawla, N. V, Lazarevic, A., Hall, L. O., & Bowyer, K. (2003). SMOTEBoost : Improving Prediction of the Minority Class in Boosting. *In European Conference on Principles of Data Mining and Knowledge Discovery (pp. 107-119). Springer, Berlin, Heidelberg.*

Chen, D., Wang, X. J., Zhou, C., & Wang, B. (2019). The Distance-Based Balancing Ensemble Method for Data With a High Imbalance Ratio. *IEEE Access*, *7*, 68940–68956. http://doi.org/10.1109/ACCESS.2019.2917920

Chen, F., & Cui, Y. (2020). Utilizing student time series behaviour in learning management systems for early prediction of course performance. *Journal of Learning Analytics*, *7*(2), 1–17. http://doi.org/10.18608/JLA.2020.72.1

Chepete, P. (2008). *Modeling of the factors affecting mathematical achievement of form 1 students in Botswana based on the 2003 Trends in International Mathematics and Science Study. Indiana University.*

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), 1–13. http://doi.org/10.1186/s12864-019-6413-7

Christian, T. M. (2014). Exploration of Classification Using NBTree for Predicting Students ' Performance. *Data and Software Engineering (ICODSE), 2014 International Conference on (pp. 1-6). IEEE.*, 0–5.

Cieslak, D. A., & Chawla, N. V. (2008). Learning decision trees for unbalanced data. *In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 241-256). Springer, Berlin, Heidelberg*, *5211 LNAI*(PART 1), 241–256. http://doi.org/10.1007/978-3-540-87479-9_34

Clark, T., & Kim, H. (2010). Use Data Mining to Improve Student Retention in Higher Education-A Case Study. *In ICEIS (1) (pp. 190-197).*

Cobbold, T. (2015). A Review of Academic Studies of Public and Private School Outcomes in Australia, (April), 1–18.

Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance.

Council, T. E. (2014). *The Most Remarkable Projects Developed by The Education Council. The Education Council.*

Del Bonifro, F., Gabbrielli, M., Lisanti, G., & Zingaro, S. P. (2020). *Student dropout prediction. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 12163 LNAI). Springer International Publishing. http://doi.org/10.1007/978-3-030-52237-7_11

Devi, G. (2011). Breast Cancer Prediction System using Feature Selection and Data Mining Methods. *International Journal of Advanced Research in Computer Science*, *2*(1), 10901–10911.

Ditzler, G., & Polikar, R. (2010). An Incremental Learning Algorithm for Non-Stationary Environments and Class Imbalance. *In Pattern Recognition (ICPR), 2010 20th International Conference on (pp. 2997-3000). IEEE.* http://doi.org/10.1109/ICPR.2010.734

Edin Osmanbegović, M. S. (2012). Data mining approach for predicting student performance. *Journal of Economics and Business, Vol. X, Issue 1, May 2012*, *X*(1), 3–12.

Education, M. of. (2011). Sultanate of Oman Muscat Declaration Conference on Education for Sustainable Development in Support of Cultural Diversity and Biodiversity Organized by the Sultanate of Oman in collaboration with the UNESCO Regional Office in Doha Muscat , Sultanate of Om, 1–7.

Education, M. of. (2013). *Education in Oman :The Drive for Quality.*

Edwin, M. D., & Sabura, F. M. (2019). Critical Analysis on Employment of Graduates in Oman. *Saudi Journal of Business and Management Studies*, *4*(7), 638–645. http://doi.org/10.21276/sjbms.2019.4.7.13

El-banna, M. (2015). A novel approach for classifying imbalance welding data : Mahalanobis genetic algorithm (MGA). *Springer-Verlag*, 407–425. http://doi.org/10.1007/s00170-014-6428-9

Ezzat, A., Wu, M., Li, X., & Kwoh, C. (2016). Drug-target interaction prediction via class imbalance-aware ensemble learning. *BMC Bioinformatics*, *17*(Suppl 19). http://doi.org/10.1186/s12859-016-1377-y

Fabich, M. (2005). A meta-analysis of demographic characteristics and learning by deaf students. *Rochester Institute of Technology.*

Farooq, M. S., Chaudhry, A. H., Shafiq, M., & Berhanu, G. (2011). Factors affecting students' quality of academic performance: A case of secondary school level. *Journal of Quality and Technology Management*, *VII*(Ii), 1–14.

Fayyad, U., Piatetsky-shapiro, G., & Smyth, P. (1996). From Data Mining to

Knowledge Discovery in. *AI Magazine*, *17*(3), 37–54.

Fern, A., & Garc, S. (2018). SMOTE for Learning from Imbalanced Data : Progress and Challenges , Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, *61*, 863–905.

Fern, A., & Jos, C. (2017). A Pareto-based Ensemble with Feature and Instance Selection for Learning from Multi-Class Imbalanced Datasets. *International Journal of Neural Systems, 27(06), 1750028.*, *27*(6), 1–21. http://doi.org/10.1142/S0129065717500289

Filges, T., Sonne-schmidt, C. S., Christian, B., & Nielsen, V. (2018). Small class sizes for improving student achievement in primary and secondary schools : a systematic review. *The Campbell Collaboration*, (February 2017).

Frenette, M., Ching, P., & Chan, W. (2015). Analytical Studies Branch Research Paper Series Academic Outcomes of Public and Private High School Students : What Lies Behind the Differences ?, (11).

Galar, M., Fern, A., Barrenechea, E., & Bustince, H. (2012). A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS*, *42*(4), 463–484.

García, V., Sánchez, J. S., & Mollineda, R. A. (2010). Exploring the performance of resampling strategies for the class imbalance problem. *In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, *6096 LNAI*(PART 1), 541–549. http://doi.org/10.1007/978-3-642-13022-9_54

Gen, B., Salahuddin, S., & Talukder, H. K. (2017). Influence of Socio-Demographic Characteristics on Academic Performance of Medical Students. *Bangladesh Journal of Medical Education*, *8*(2), 18–23.

Ghorbani, R., & Ghousi, R. (2020). Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques. *IEEE Access*, *8*, 67899–67911. http://doi.org/10.1109/ACCESS.2020.2986809

Gu, Q., Cai, Z., Zhu, L., & Huang, B. (2008). Data Mining on Imbalanced Data Sets. *In Advanced Computer Theory and Engineering, 2008. ICACTE'08. International Conference on (pp. 1020-1024). IEEE.*, 1020–1024. http://doi.org/10.1109/ICACTE.2008.26

Guo, H. (2004). Learning from Imbalanced Data Sets with Boosting and Data Generation : The DataBoost-IM Approach. *ACM Sigkdd Explorations Newsletter*, *6*(1), 30–39.

Guo, H., Liu, H., Wu, C., Zhi, W., Xiao, Y., & She, W. (2016). Logistic

discrimination based on G-mean and F-measure for imbalanced problem. *Journal of Intelligent & Fuzzy Systems*, *31*, 1155–1166. http://doi.org/10.3233/IFS-162150

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining : Concepts and Techniques : Concepts and Techniques (3rd Edition). Data Mining*. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/B9780123814791000010

Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K. U., & Sattar, M. U. (2020). Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Applied Sciences (Switzerland)*, *10*(11). http://doi.org/10.3390/app10113894

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on (pp. 1322-1328). IEEE.*, (3), 1322–1328.

He, H., & Garcia, E. A. (2009a). Learning from Imbalanced Data. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 21, NO. 9*, *21*(9), 1263–1284.

He, H., & Garcia, E. A. (2009b). Learning from Imbalanced Data. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, *21*(9), 1263–1284.

Herzi, A. A. (2010). Development Education in the Era of globalization : A Case from Malaysia and Oman higher education Sector, 1–17.

Hlosta, M., Stríž, R., Kupčík, J., Zendulka, J., & Hruška, T. (2013). Constrained Classification of Large Imbalanced Data by Logistic Regression and Genetic Algorithm. *International Journal of Machine Learning and Computing*, *3*(2), 214–218. http://doi.org/10.7763/ijmlc.2013.v3.305

Hu, S., Liang, Y., Ma, L., & He, Y. (2009). MSMOTE: Improving classification performance when training data is imbalanced. *2nd International Workshop on Computer Science and Engineering, WCSE 2009*, *2*, 13–17. http://doi.org/10.1109/WCSE.2009.756

Hulse, J. Van, Khoshgoftaar, T. M., & Napolitano, A. (2011). An exploration of learning when data is noisy and imbalanced. *Intelligent Data Analysis*, *15*, 215–236. http://doi.org/10.3233/IDA-2010-0464

Hung, J., & Zhang, K. (2008). Revealing Online Learning Behaviors and Activity Patterns and Making Predictions with Data Mining Techniques in Online Teaching. *MERLOT Journal of Online Learning and Teaching.*, *4*.

Hussain, S., Muhsin, Z. F., Salal, Y. K., Theodorou, P., Kurtoğlu, F., & Hazarika, G.

C. (2019). Prediction model on student performance based on internal assessment using deep learning. *International Journal of Emerging Technologies in Learning*, *14*(8), 4–22. http://doi.org/10.3991/ijet.v14i08.10001

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem : A systematic study. *Intelligent Data Analysis*, *6*, 429–449.

Jayaraman, J. D. (2020). Predicting Student Dropout by Mining Advisor Notes. *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, (Edm), 629–632.

Jedrzejowicz, J., Kostrzewski, R., Neumann, J., & Zakrzewska, M. (2018). Imbalanced data classification using MapReduce and relief. *Journal of Information and Telecommunication*, *1839*, 1–14. http://doi.org/10.1080/24751839.2018.1440454

Johannes, N. (2017). The relationship between demographics and the academic achievement of engineering students. *3rd International Conference on Higher Education Advances*, 347–355.

Johnson, R. A., & Chawla, N. V. (2016). *Data Science for Imbalanced Data: Methods and Applications*.

Jones, K. R., & Ezeife, A. N. (2011). School Size as a Factor in the Academic Achievement of Elementary School Students. *Psychology*, *2*(8), 859–868. http://doi.org/10.4236/psych.2011.28131

Kabakchieva, D. (2012). Student Performance Prediction by Using Data Mining Classification Algorithms. *International Journal of Computer Science and Management Research*, *1*(4), 686–690.

Kabakchieva, D. (2013). Predicting Student Performance by Using Data Mining Methods for Classification Dorina Kabakchieva. *Cybernetics and Information Technologies, 13(1), 61-72.*, *13*(1), 61–72. http://doi.org/10.2478/cait-2013-0006

Kaitlin, K., Smith, T., & Sadler, B. (2018). Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets. *Data Science Review*, *1*(3), 9. Retrieved from https://scholar.smu.edu/datasciencereviewhttp://digitalrepository.smu.edu.Availableat:https://scholar.smu.edu/datasciencereview/vol1/iss3/9

Kamal, M., & Bener, A. (2009). Factors Contributing to School Failure among School Children in a very Fast Developing Arabian Society. *Oman Medical Journal*, *24*(3), 212–217. http://doi.org/10.5001/omj.2009.42

Kaur, A., Kaur, K., & Jain, S. (2016). Predicting Software Change-Proneness with Code Smells and Class Imbalance Learning. *In Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on*

(pp. 746-754). IEEE., 746–754.

Kee, A. mohd remaliMohamad A. G. K. K. Y. (2013). Understanding Academic Performance Based On Demographic Factors, Motivation Factors & Learning Styles. *International Journal of Asian Social Science*, *3*(9), 1938–1951.

Kermanidis, K., Maragoudakis, M., Fakotakis, N., & Kokkinakis, G. (2002). Learning Greek Verb Complements : Addressing the Class Imbalance, (Laurikkala 2001).

Khalilia, M., Chakraborty, S., & Popescu, M. (2011). Predicting disease risks from highly imbalanced data using random forest. *BMC Medical Informatics and Decision Making*, *11*(1), 51. http://doi.org/10.1186/1472-6947-11-51

Khoshgoftaar, T. M., Seliya, N., & Drown, D. J. (2010). Evolutionary data analysis for the class imbalance problem. *Intelligent Data Analysis*, *14*, 69–88. http://doi.org/10.3233/IDA-2010-0409

Konstantopoulos, S. (2005). Trends of School Effects on Student Achievement : Evidence from NLS : 72 , HSB : 82 , and NELS : 92, (1749).

Koutina, M., Kermanidis, K., Koutina, M., Kermanidis, K., Postgraduate, P., Performance, S., & Machine, U. (2017). Predicting Postgraduate Students ' Performance Using Machine Learning Techniques. *Artificial Intelligence Applications and Innovations*.

Kovačić, Z. (2010). Early Prediction of Student Success: Mining Students Enrolment Data. *Proceedings of Informing Science & IT Education Conference*, 647–665.

Krawczyk, B. (2016). Learning from imbalanced data : open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221–232. http://doi.org/10.1007/s13748-016-0094-0

Kretzschmar, R., & Eggimann, F. (2005). Feedforward neural network models for handling class overlap and class imbalance. *International Journal of Neural Systems*, *15*(5), 323–338.

Kubus, M. (2020). Evaluation of resampling methods in the class unbalance problem. *Econometrics*, *24*(1), 39–50. http://doi.org/10.15611/eada.2020.1.04

Kumar, M. (2017). Literature Survey on Student ' s Performance Prediction in Education using Literature Survey on Student ' s Performance Prediction in Education using Data Mining Techniques. *International Journal of Education and Management Engineering*, (October). http://doi.org/10.5815/ijeme.2017.06.05

Lee, S., & Chung, J. Y. (2019). The machine learning-based dropout early warning system for improving the performance of dropout prediction. *Applied Sciences (Switzerland)*, *9*(15). http://doi.org/10.3390/app9153093

Lema, G., Nogueira, F., West, W. S., Mv, O., & Aridas, C. K. (2017). Imbalanced-learn : A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *The Journal of Machine Learning Research, 18(1), 559-563, 18*, 1–5.

Li, J., Fong, S., Sung, Y., Cho, K., Wong, R., & Wong, K. K. L. (2016). Adaptive swarm cluster-based dynamic multi-objective synthetic minority oversampling technique algorithm for tackling binary imbalanced datasets in biomedical data classification. *BioData Mining*, 1–15. http://doi.org/10.1186/s13040-016-0117-1

Lim, T. W., Khor, K. C., & Ng, K. H. (2019). Dimensionality reduction for predicting student performance in unbalanced data sets. *International Journal of Advances in Soft Computing and Its Applications*, *11*(2), 76–86.

Lin, W., & Chen, J. J. (2012). Class-imbalanced classifiers for high-dimensional data. *BRIEFINGS IN BIOINFORMATICS*, *14*(1). http://doi.org/10.1093/bib/bbs006

Longadge & Dongre. (2013). Class Imbalance Problem in Data Mining : Review. *International Journal of Computer Science and Network (IJCSN)*, *2*(1).

López, V., Fernández, A., Moreno-torres, J. G., & Herrera, F. (2012). Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics. *Expert Systems with Applications*, *39*, 6585–6608. http://doi.org/10.1016/j.eswa.2011.12.043

Lubienski, C. (2006). Charter , Private , Public Schools and Academic Achievement : New Evidence from NAEP Mathematics Data 1. *National Center for the Study of Privatization in Education, Teachers College, Columbia University*, *16*.

Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216–231. http://doi.org/10.1016/j.patcog.2019.02.023

M.Mustapha, Md.Nasir, Z. M. (2010). ranking of influence factors in predicting student academic performance. *Information Technology Journal 9*.

Maciejewski, T., & Stefanowski, J. (2011). Local Neighbourhood Extension of SMOTE for Mining Imbalanced Data. *Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on (pp. 104-111). IEEE.*

Maheshwari, S., Agrawal, P. J., & Sharma, S. (2011). A New approach for Classification of Highly Imbalanced Datasets using Evolutionary Algorithms. *International Journal of Scientific & Engineering Research*, *2*(7), 1–5.

Martino, D., Decia, F., Molinelli, J., & Fern, A. (2010). Improving Electric Fraud Detection using Class Imbalance Strategies. *In ICPRAM (2)*, 135–141.

Matveev, K. (2013). Education in Oman The Drive for Quality. *The World Bank*.

Maurya, Chandresh Kumar, D. T., & Vijendran, G. (2017). Distributed Sparse Class-Imbalance Learning and its Applications. *IEEE Transactions on Big Data.*, *13*(9). http://doi.org/10.1109/TBDATA.2017.2688372

Mduma, N., Kalegele, K., & Machuve, D. (2019). Machine learning approach for reducing students dropout rates. *International Journal of Advanced Computer Research*, *9*(42), 156–169. http://doi.org/10.19101/ijacr.2018.839045

Mirza, B., Lin, Z., Cao, J., & Lai, X. (2015). Voting based Weighted Online Sequential Extreme Learning Machine for Imbalance Multi-Class Classification. *In 2015 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 565-568). IEEE.*, 565–568.

Mishra, T. (2014). Mining Students ' Data for Performance Prediction. *2014 Fourth International Conference on Advanced Computing & Communication Technologies*, 255–262. http://doi.org/10.1109/ACCT.2014.105

MOE. (2009). *ICT and Education in the Sultanate of Oman.*

Mohamed, A., Husain, W., & Rashid, A. (2015). Review on Predicting Student ' s Performance using Data Mining Techniques. *Procedia Computer Science*, *72*, 414–422. http://doi.org/10.1016/j.procs.2015.12.157

Mosoumeh Zareapoo, P. S. (2015). Application of Credit Card Fraud Detection: Based on Bagging Esaemble Classifier. *Elsevier B V By-Nc-Nd, C C Conference, International*, *48*(Iccc), 679–685. http://doi.org/10.1016/j.procs.2015.04.201

Mullis. (2015). *TIMSS 2015 International Results in Mathematics.*

Muthurman, S., Veerasamy, R., & Al-Hazaizi, M. (2020). E-learning to enhance educational competitiveness in the sultanate of Oman. *International Journal of Innovation, Creativity and Change*, *11*(2), 84–92.

Najar, N. Al. (2016). View of education development in Oman. *International Journal of Academic Research in Education and Review, 4(1), 10-18.*, *7*(March), 133–145.

Nasir, M. (2012). Demographic characteristics as correlates of academic achievement of university students. *Academic Research International*, *2*(2), 400–405.

Nayak, S. A. K., & Krishna, R. (2018). Comparing the Behavior of Oversampling and Undersampling Approach of Class Imbalance Learning by Combining Class Imbalance Problem with Noise. *In ICT Based Innovations (pp. 23-30). Springer, Singapore.*, *653*(January). http://doi.org/10.1007/978-981-10-6602-3

NCSIOman. (2017). Population Projections. *The Lancet*, *296*(7672), 563. http://doi.org/10.1016/S0140-6736(70)91362-0

Ofelia, M., Pedro, Z. S., Baker, R. S. J., Bowers, A. J., & Heffernan, N. T. (2013).

Predicting College Enrollment from Student Interaction with an Intelligent Tutoring System in Middle School. *Educational Data Mining 2013*.

Olson, D. L., & Delen, D. 2 Data Mining Process, Springer (2008).

Onan, A. (2019). Consensus Clustering-Based Undersampling Approach to Imbalanced Learning. *Scientific Programming*, *2019*. http://doi.org/10.1155/2019/5901087

Onoyase. (2015). Academic Performance Among Students In Urban , Semi- Urban And Rural Secondary Schools Counselling Implications. *Developing Country Studies*, *5*(19), 122–126.

Ortigosa-Hernández, J., Inza, I., & Lozano, J. A. (2017). Measuring the class-imbalance extent of multi-class problems. *Pattern Recognition Letters*, *98*, 32–38. http://doi.org/10.1016/j.patrec.2017.08.002

Osman, D. M. E. T. (2010). Educational Portal in Oman: Toward a connected community. *Journal of American Arabic Academy for Sciences and Technology*.

Pal, A. K. (2013). Analysis and Mining of Educational Data for Predicting the Performance of Students. *International Journal of Electronics Communication and Computer Engineering*, *4*(5), 1560–1565.

Paola, M. De, Scoppa, V., & Ponzo, M. (2013). Class size effects on student achievement: heterogeneity across abilities and fields. *Education Economics*, (August). http://doi.org/10.1080/09645292.2010.511811

Pardos, Z. A., & Heffernan, N. T. (2010). Using HMMs and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset. *Journal of Machine Learning Research W & CP*.

Pelayo, L., & Dick, S. (2007). Applying novel resampling strategies to software defect prediction. *Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS*, 69–72. http://doi.org/10.1109/NAFIPS.2007.383813

Peng, Y., & Yao, J. (2010). AdaOUBoost : Adaptive Over-sampling and Under-sampling to Boost the Concept Learning in Large Scale Imbalanced Data Sets. *In Proceedings of the International Conference on Multimedia Information Retrieval (pp. 111-118). ACM.*, 111–118.

Phua, C., Alahakoon, D., & Lee, V. (2006). Minority Report in Fraud Detection : Classification of Skewed Data. *ACM Sigkdd Explorations Newsletter*, *6*(1), 50–59.

Phung, S. L. (2009). Learning pattern classification tasks with imbalanced data sets (pp. 193–208).

Rahman, M. M., & Davis, D. N. (2013). Cluster based under-sampling for unbalanced cardiovascular data. *In Proceedings of the World Congress on Engineering, 3 LNECS*, 1480–1485.

Ramesh, V. P. P. K. R. (2014). Predicting Student Performance : A Statistical and Data Mining Approach. *International Journal of Computer Applications*, (February 2013), 34–39.

Razzaghi, T., Roderick, O., Safro, I., & Marko, N. (2016). Multilevel Weighted Support Vector Machine for Classification on Healthcare Data with Missing Values. *PLOS ONE*, 1–19. http://doi.org/10.1371/journal.pone.0155119

Road, W. (2014). Fault detection for the class imbalance problem in semiconductor ¤ manufacturing processes. *Journal of Circuits, Systems, and Computers*, *23*(4), 1–20. http://doi.org/10.1142/S0218126614500492

Rodríguez, E., Arqués, J. L., Rodríguez, R., Nuñez, M., Medina, M., Talarico, T. L., … Masuelli, M. (2019). Educational reform in oman system and structural changes. In *Intech* (Vol. 32, pp. 137–144). Retrieved from https://www.intechopen.com/books/advanced-biometric-technologies/liveness-detection-in-biometrics

Sánchez-Hernández, F., Ballesteros-Herráez, J. C., Kraiem, M. S., Sánchez-Barba, M., & Moreno-García, M. N. (2019). Predictive modeling of ICU healthcare-associated infections from imbalanced data. Using ensembles and a clustering-based undersampling approach. *Applied Sciences (Switzerland)*, *9*(24). http://doi.org/10.3390/app9245287

Saravanan, N., & Gayathri, V. (2018). Performance and Classification Evaluation of J48 Algorithm and Kendall's Based J48 Algorithm (KNJ48). *International Journal of Computer Trends and Technology*, *59*(2), 73–80. http://doi.org/10.14445/22312803/ijctt-v59p112

Sarrayrih, M. A., & Sriram, B. (2015). Major challenges in developing a successful e-government : A review on the Sultanate of Oman. *Journal of King Saud University - Computer and Information Sciences*, *27*(2), 230–235. http://doi.org/10.1016/j.jksuci.2014.04.004

Sathic Ali, P. U., & Ventakeswaran, C. J. (2011). Improved Evidence Theoretic kNN Classifier based on Theory of Evidence. *International Journal of Computer Applications*, *15*(5), 37–41. http://doi.org/10.5120/1943-2597

Seiffert, C., Khoshgoftaar, T. M., & Hulse, J. Van. (2009). Hybrid sampling for imbalanced data. *Integrated Computer-Aided Engineering*, *16*, 193–210. http://doi.org/10.3233/ICA-2009-0314

Seiffert, C., Khoshgoftaar, T. M., Hulse, J. Van, & Napolitano, A. (2010). RUSBoost : A Hybrid Approach to Alleviating Class Imbalance. *IEEE*

TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, *40*(1), 185–197.

Shelke, M. S., Deshmukh, P. R., & Shandilya, P. V. K. (2017). A Review on Imbalanced Data Handling Using Undersampling and Oversampling Technique. *International Journal of Recent Trends in Engineering & Research (IJRTER)*, 444–449.

Shrive, F. M., Stuart, H., Quan, H., & Ghali, W. A. (2006). Dealing with missing data in a multi-question depression scale: A comparison of imputation methods. *BMC Medical Research Methodology*, *6*, 1–10. http://doi.org/10.1186/1471-2288-6-57

Sohajbir Singh Ubha, G. K. B. (2016). Data Mining for Prediction of Students' Performance in the Secondary Schools of the State of Punjab. *International Journal of Innovative Research in Computer and Communication Engineering*, 15339–15346. http://doi.org/10.15680/IJIRCCE.2016.

Solomon, D. (2018). Predicting Performance and Potential Difficulties of University Student using Classification : Survey Paper. *International Journal of Pure and Applied Mathematics*, *118*(18), 2703–2707.

Soonthornphisaj, N., Sira-Aksorn, T., & Suksankawanich, P. (2018). Social Media Comment Management using SMOTE and Random Forest Algorithms. *International Journal of Computational Intelligence Systems*, *8*(2), 54–58. http://doi.org/10.1080/XXXXXXXXXXXXXX

Statistics, N. C. for E. (2017). *Highlights From TIMSS and TIMSS Advanced 2015*.

Stefanowski, J., & Wilk, S. (2006). Rough Sets for Handling Imbalanced Data : Combining Filtering and Rule-based Classifiers. *Fundamenta Informaticae*, *72*, 379–391.

Su, P., Mao, W., Zeng, D., Li, X., & Wang, F. (2009). Handling Class Imbalance Problem in Cultural Modeling. *In Intelligence and Security Informatics, 2009. ISI '09. IEEE International Conference on (pp. 251-256). IEEE.*, 251–256.

Sun, Y., Wong, A. K., & Kamel, M. S. (2009). Classification of imbalanced data : a review c World Scientific Publishing Company. *International Journal of Pattern Recognition and Artificial Intelligence*, (November). http://doi.org/10.1142/S0218001409007326

Sundar, P. (2015). A Comparative Study to Predict Student's Performance Using Bayesian Network. *IOSR Journal of Engineering (IOSRJEN)*, (March 2013), 36–42. http://doi.org/10.9790/3021-03213742

Superby, J. F. (2006). Determination of factors influencing the achievement of the first-year university students using data mining methods. *Workshop on Educational Data Mining (Vol. 32, P. 234)*.

Tang, Y., & Zhang, Y. (2009). SVMs Modeling for Highly Imbalanced Classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39(1), 281-288*, *39*(1), 281–288.

Thai-nghe, N., Busche, A., & Schmidt-thieme, L. (2009). Improving Academic Performance Prediction by Dealing with Class Imbalance. *In Intelligent Systems Design and Applications, 2009. ISDA'09. Ninth International Conference on (pp. 878-883). IEEE.*

Thakar, P., Mehta, A., & Manisha. (2017). A unified model of clustering and classification to improve students' employability prediction. *International Journal of Intelligent Systems and Applications*, *9*(9), 10–18. http://doi.org/10.5815/ijisa.2017.09.02

Thammasiri, D., Delen, D., Meesad, P., & Kasap, N. (2013). A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *EXPERT SYSTEMS WITH APPLICATIONS*. http://doi.org/10.1016/j.eswa.2013.07.046

Thiele, T., Pope, D., Singleton, A., & Stanistreet, D. (2016). Role of students ' context in predicting academic performance at a medical school : a retrospective cohort study. *BMJ Open*, 1–11. http://doi.org/10.1136/bmjopen-2015-010169

Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2020). Transfer learning from deep neural networks for predicting student performance. *Applied Sciences (Switzerland)*, *10*(6). http://doi.org/10.3390/app10062145

Ustyannie, W., & Suprapto, S. (2020). Oversampling Method To Handling Imbalanced Datasets Problem in Binary Logistic Regression Algorithm. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *14*(1), 1. http://doi.org/10.22146/ijccs.37415

Utzman, R. R., Riddle, D. L., & Jewell, D. V. (2007). Use of Demographic and Quantitative. *Physical Therapy*, *87*(9).

Uyeno, R. (2006). The Role of Demographic Factors in Predicting Student Performance on a State Reading Test. *Online Submission*.

Verma, A. (2019a). Evaluation of Classification Algorithms with Solutions to Class Imbalance Problem on Bank Marketing Datas ... *International Research Journal of Engineering and Technology*, 54–60. Retrieved from https://www.academia.edu/download/59954293/IRJET-V6I30820190707-55305-1e47yyf.pdf

Verma, A. (2019b). Evaluation of Classification Algorithms with Solutions to Class Imbalance Problem on Bank Marketing Datas ... *International Research Journal of Engineering and Technology*, 54–60.

Visa, S. (2005). Issues in mining imbalanced data sets-a review paper. *In Proceedings of the Sixteen Midwest Artificial Intelligence and Cognitive Science Conference (Vol. 2005, Pp. 67-73). Sn.*

Wallace, B. C., Small, K., Brodley, C. E., & Trikalinos, T. A. (2011). Class Imbalance , Redux. *In Data Mining (ICDM), 2011 IEEE 11th International Conference on (pp. 754-763). IEEE.* http://doi.org/10.1109/ICDM.2011.33

Wang, S. (2011). *Ensemble diversity for class imbalance learning.*

Wang, S., Minku, L. L., & Yao, X. (2018). A Systematic Study of Online Class Imbalance Learning With Concept Drift. *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, 1–20.

Wang, S., Minku, L. L., & Yao, X. I. N. (2013). Online class imbalance learning and its applications in fault detection. *International Journal of Computational Intelligence and Applications, 12(04), 1340001.*, *12*(4), 1–19. http://doi.org/10.1142/S1469026813400014

Wang, S., & Yao, X. (2009). Diversity Analysis on Imbalanced Data Sets by Using Ensemble Models. *In Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on (pp. 324-331). IEEE.*

Wang, S., & Yao, X. (2013). Using Class Imbalance Learning for Software Defect Prediction. *IEEE Transactions on Reliability*, *62*(2), 434–443.

Weng, C. G., & Poon, J. (2008). A new evaluation measure for imbalanced datasets. *Conferences in Research and Practice in Information Technology Series*, *87*(81373883), 27–32.

Wolff, A., Zdrahal, Z., & Pantucek, M. (2013). Improving retention : predicting at-risk students by analysing clicking behaviour in a virtual learning environment. *Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 145-149). ACM.*, 145–149.

Yazdi, J. S., Kalantary, F., & Yazdi, H. S. (2013). Investigation on the Effect of Data Imbalance on Prediction of Liquefaction. *INTERNATIONAL JOURNAL OF GEOMECHANICS*, *1*(AUGUST), 463–467. http://doi.org/10.1061/(ASCE)GM.1943-5622.0000217.

Yo, K. (2012). Mining Educational Data to Improve Students ' Performance : A Case Study Mining Educational Data t o Improve Students ' Performance : A Case Study. *researchgate.net/publication/258501044*, (January).

Yu, H., Ni, J., Xu, S., Qin, B., & Jv, H. (2014). Estimating harmfulness of class imbalance by scatter matrix based class separability measure. *Intelligent Data Analysis*, *18*(2), 203–216. http://doi.org/10.3233/IDA-140637

Zafra, A., & Ventura, S. (2009). Predicting Student Grades in Learning Management

Systems with Multiple Instance Genetic Programming. *International Working Group on Educational Data Mining.*, (Mil), 307–314.

Zhang, C. (2017). Feature Selection and Resampling in Class Imbalance Learning : Which Comes First ? An Empirical Study in the Biological Domain. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM).*

Zhang, M., & Li, Y. (2015). Towards Class-Imbalance Aware Multi-Label Learning. *International Joint Conference on Artificial Intelligence (IJCAI 2015)*, (Ijcai), 4041–4047.

Zhao, Y., Wong, Z. S., & Tsui, K. L. (2018). A Framework of Rebalancing Imbalanced Healthcare Data for Rare Events ' Classification : A Case of Look-Alike Sound-Alike Mix-Up Incident Detection. *Journal of Healthcare Engineering*, *2018*(2010).

Zhi, W., Guo, H., Fan, M., & Ye, Y. (2015). Instance-based ensemble pruning for imbalanced learning. *Intelligent Data Analysis*, *19*, 779–794. http://doi.org/10.3233/IDA-150745

Zhou, Z., Member, S., & Liu, X. (2006). Training Cost-Sensitive Neural Networks with Methods Addressing the Class Imbalance Problem. *IEEE Transactions on Knowledge and Data Engineering*, *18*(1), 63–77.

Zhu, J., Hovy, E., & Rey, M. (2007). Active Learning for Word Sense Disambiguation with Methods for Addressing the Class Imbalance Problem. *In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL).*, (June), 783–790.