Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Data Mining Techniques to Improve Predictions Accuracy of Students' Academic Performance: A Case Study with Xorro-Q

A thesis presented in a partial fulfilment of the requirements for
Master of Information Science (IT)
At
Massey University
Auckland
New Zealand
in 2018

Gomathy Suganya
(Supervisor: Dr. Teo Susnjak)
(Co-Supervisor: Dr.Anuradha Mathrani)

Abstract

Recent research in analytics has assisted policy makers capitalize on their ever-increasing data repositories and make data-driven predictions to create a vision for developing strategies to achieve their business targets. This is especially relevant in educational environments where data mining techniques can be applied to make predictions around students' academic performance. This can help educators align a teaching strategy which encourages and assists students with their learning. Suitable pedagogical support can be provided to enhance the overall student learning experience.

This study is in the educational domain where student-related course data has been used to extract insights on student performances over the study period. Extensive data collected from an educational tool (Xorro-Q) used in an engineering course delivery has aided this investigation. Data collected from Xorro-Q comprised student scores from real-time and self-paced activities set by educators over a 12-week semester period along with students' final Exam scores and scores from a compulsory prerequisite course. Popular data mining techniques have been applied to predict the academic performance of students based on data extracted from Xorro-Q. This is done by training the classifier using four different algorithms, namely, Naive Bayes, Logistic regression, K nearest neighbour and Random Forest. Process mining techniques have been applied along with the general features to find out the effectiveness, such as improvement in accuracy of predictions. The study has further implications in enhancing value of the role of analytics for predictive modelling by incorporating process mining features in the training set of data.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Dr. Teo Susnjak and Co-supervisor Dr. Anuradha Mathrani for providing valuable guidance, comments and suggestions throughout the course of my project. But for their help I would not have completed my project successfully.

I would like to acknowledge my client Pablo Garcia for providing me the data and for the countless hours he has spent in discussing with me the way forward and the potential outcomes of this project.

I wish to express my sincere thanks to Dr.James Lim for clarifying my doubts and giving valuable suggestions for correction modification and improvement of the project.

Finally Iam thankful and fortunate enough to get unconditional love and constant encouragement from my friend Rahila Umer and my little son Sai Dhanwin which helped me in successfully completing the project on time.

PUBLICATIONS AND PRESENTATIONS

Publications generated from this project so far:

Suganya, G., Susnjak, T., Mathrani, A., Lim, J., Garcia, P. (2017). Data Mining Techniques to Improve Predictions Accuracy of students' Academic Performance: A Case Study with Xorro-Q. *Proceedings of the 11th International conference on Data Mining. Computers, Communication and Industrial Applications.*

Presentations given from this project so far:

1. Suganya, G., Susnjak, T., Mathrani, A., Lim, J., Garcia, P. (2017). Data Mining Techniques to Improve Predictions Accuracy of Students' Academic Performance: A Case Study with Xorro-Q. *Oral presentation at the 11th International conference on Data Mining. Computers, Communication and Industrial Applications, Kuala Lumpur, Malaysia.*

2. Suganya, G., Susnjak, T., Mathrani, A., Lim, J., Garcia, P. (2017). Data Mining Techniques to Improve Predictions Accuracy of Students' Academic Performance: A Case Study with Xorro-Q. *Poster session presented at the 5th INMS Postgraduate students at Massey University, Auckland, New Zealand, 26th October 2017.*

3.Suganya, G.(2017).*Data Mining Techniques to Improve Predictions Accuracy of Students' Academic Performance: A Case Study with Xorro-Q.* Seminar delivered on 25th August 2017 to academic and industry personnel involved in this project.

ACHIEVEMENTS

Award won Suganya, G., Susnjak, T., Mathrani, A., Lim, J., Garcia, P. (2017). Data Mining Techniques to Improve Predictions Accuracy of Students' Academic Performance: A Case Study with Xorro-Q. Won the best session paper in *11th International Conference on Data Mining, Computers, Communication and Industrial Applications.*

Contents

List of Figures

List of Tables