Employing data to enhance teaching and learning in MENDAKI Tuition Scheme (MTS)

Siti Afiyah Mustapha
Yayasan MENDAKI
Singapore

JiZhi Li
Singapore University
of Social Sciences
Singapore

Sylvia Chong
Singapore University
of Social Sciences
Singapore

The Malay community in Singapore has made significant social and economic progress over the years. In 2018, Singapore's Ministry of Education released figures reflecting educational achievements over a 10-year period between 2008 and 2017. Indicators such as progress in national examinations and increased representation of Malay students in post-secondary institutions over the last decade are encouraging milestones of educational mobility. The community's educational performance for key national examinations however is generally lower than that of other ethnic groups. Several researchers have highlighted the "under-attainment" of Singapore Malays. The MENDAKI Tuition Scheme (MTS), MENDAKI's flagship programme, provides highly subsidised tuition to support socioeconomically disadvantaged students in their learning. This paper reports the pilot phase of a collaboration between a university in Singapore and MENDAKI. The primary aim of the collaboration is to leverage on the use analytics to enhance the teaching and learning of disadvantaged Malay students in the MTS. Other secondary aims of the collaboration are to establish structures and to build capability of MENDAKI in the use of data for evidence-based decision making.

Keywords: Malays, MENDAKI, low performing community, analytics

Context of Project

The Malay community in Singapore has made significant social and economic progress over the years. In 2018, the education statistics from over a 10-year period between 2008 and 2017 were released (Ministry of Education, 2018). Indicators such as progress in national examinations and increased representation of Malay students in post-secondary institutions are encouraging milestones of educational mobility. The proportion of Malay primary one pupils who move on to post-secondary education has doubled from 45 per cent in 1995 to 94 per cent in 2017, whereas those who eventually obtained degrees and diplomas have increased from 15 per cent in 2010 to 21 per cent in 2015 (Toh, 2017). Only 1 per cent of its children do not complete 10 years of schooling (Channel NewsAsia, 2018). The percentage of professionals, managers, executives and technicians (PMETs) from the community has also jumped from slightly more than 7 per cent in 1980 to more than 32 per cent in 2015 (Channel NewsAsia, 2018). Figure 1 shows the percentage of Malay students who further their education and go on to post-secondary educational institutions such as Junior Colleges (JC), Polytechnics, Institutes of Technical Education (ITE) and other private institutions.

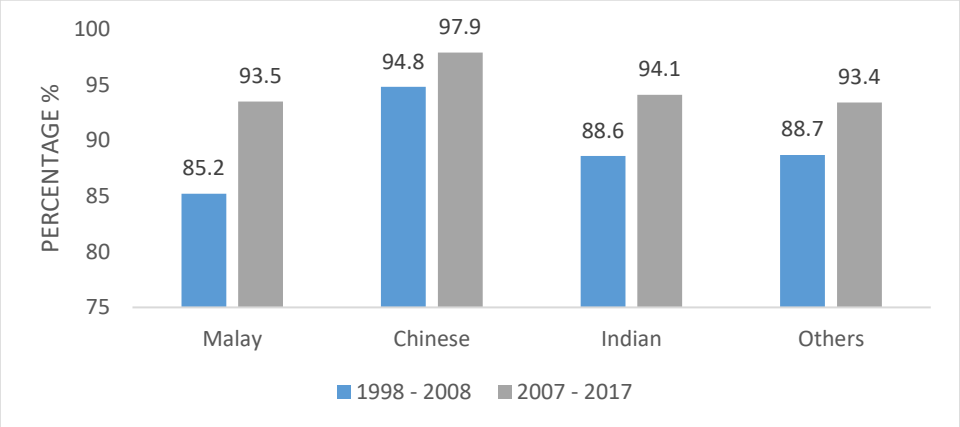

Figure 1: Percentage of P1 Cohort That Progressed To Post-Secondary Education (MOE Statistics Digest, 2018)

In relative terms however, the Malay community's educational performance for key national examinations is generally lower than that of other ethnic groups. Several researchers have highlighted the "under-attainment" of the Malay community when compared with other ethnic communities in Singapore (Tan and Ho, 2001, Zhang, 2014).

In 1982, Yayasan MENDAKI was established 'to empower the community through excellence in education' (Yayasan MENDAKI, 2017). MENDAKI programmes are designed to supplement or complement national education initiatives (Yayasan MENDAKI, 2017). The MENDAKI Tuition Scheme (MTS) provides highly-subsidised tuition to help Malay and/or Muslim students attain better results in their school and national examinations. The flagship program started off with only 880 students in 1982. Today, more than 9,000 students enrol in the programme annually. It runs over 50 centres for students from Primary 1 to Secondary 5. Students in MTS come largely from disadvantaged background. Over the years, MTS has been a valuable tool to improve Malay academic underperformance and to help Malay-Muslims acquire a higher chance of social mobility. Beyond provision of academic support, MTS empowers disadvantaged families with resources and educators with pedagogical techniques to engage students towards school success and to cultivate positive life-long learners' identity.

Along with these expansions is a rapidly growing source of student information/data to be mined for insightful knowledge. Having access to a rich data source potentially enhances the ability to effectively identify, analyse and provide interventions for low SES and/or low performing students. This paper reports the pilot phase of a collaboration between a Singapore university and MENDAKI. The pilot phase profiled of a group of 875 MTS students as well as identify determinants of their academic performance to explore the use of analytics to generate in-depth information for data-driven and evidence-based decision making.

Pilot Phase

The objective of conducting a pilot study was to examine the feasibility of a data mining approach that is intended to be used in a larger scale study. The research team explored and examined a MENDAKI data set to transform the data into a format and nature that is suitable for modelling using analytics tools. The pilot was also a knowledge discovery process that generated meaningful insights to the MTS students. This was important in shaping the research protocol of the larger study.

Sample Population

In 2015 there were 10,005 students enrolled in MTS. This comprised of 6,060 (~60%) primary school students and 3,945 (~40%) secondary school students. Primary school students may be taking Foundation or Standard subjects or a mixture of both, while Secondary level students were from Normal Technical, Normal Academic or Express stream, some may be taking subjects from a different stream. For the pilot phase, data variables of 875 Primary 6 students who sat for the Primary School Leaving Examinations (PSLE) standard subjects were analysed.

To establish a common performance indicator for benchmarking the academic performance of this pilot group of MTS students, a standardised measurement was essential for deriving fair and valid conclusion. In this dataset, academic variables were either school-based exams or PSLE results. In view of the varying standards of school-based exams, the PSLE results were used as target variable.

Methodology

This study applied K-means and Model-based Clustering methods to explore the characteristics of different groupings within this dataset. Cluster analysis is the grouping of similar objects together to form "clusters" where the objects within each have similar traits and characteristics. If the sample population converges to similar clustering rules, the "natural" groupings of the sample are established. The idea of similarity can be defined differently depending on the focus of the analysis, common measures include Euclidean distance and correlation coefficient. Unlike supervised learning methods such as regression models and decision trees where models are constructed using a target variable to make prediction, cluster analysis is unsupervised, which means the interest is solely on the input variables and interpretation of results can be rather subjective.

The algorithms of K-means and Model-based Clustering form the two major groups of clustering algorithms. K-means Clustering uses the heuristic method, which is more intuitive as it measures similarity using high-dimensional Euclidean distance between data points. Model-based Clustering is non-heuristic, as it consists of a

formal statistical structure and model where each data point has a probability of belonging to each cluster. The analysis is a holistic inference of the cluster mean, variance as well as mathematical formula for prediction.

To establish a number of clusters which balances both statistical optimum and practicality, the two clustering algorithms are tested on two-cluster and three-cluster scenario to compare the results. Despite slight deviation in the distributions of variables, K-means and Model-based Clustering produce near similar results. Model-based Clustering was selected for the final analysis as it can better manage a range of different variables (nominal, ordinal etc.) through formal modelling and allows for comprehensive understanding of the cluster structures. Between the two models, the two-clusters model has more clear-cut characteristics whereas the three-clusters model produced some ambiguous or overlapping traits and one of its cluster behave more like a residual cluster rather than a meaningful group. The final model used for analysis will be a two-clusters model derived using Model-based Clustering algorithm.

Input Variables

A total of 131 variables were in the dataset. This include students' demographic, academic, socioeconomic-status (SES) as well as MTS information. Demographic variables include fundamental information of the student and his/her family composition. Academic variables include school-based and national exam results. SES variables include family income, subsidies and housing information. MTS variables include course enrolment information.

The data preparation included cleaning up the variable selection. Aside from removing irrelevant variables, new variables were derived for better representation of student population. Variables recoded into numeric values so that they can be used in data mining algorithms. The variables can be grouped into four categories:

- Academic: PSLE aggregate
- Demographic: gender, race, family role of student, marital status of main applicant, number of household members, number of other children in the household enrolled in MENDAKI, number of household members aged 65 and above and number of children in the household aged 6 and below
- Socioeconomic status: type of residence, rental block status, Per Capita Income (PCI)
- MTS: programme

The PSLE Aggregate is placed into three categories of eligible Secondary School streams. This was based on national cut-offs for Normal Technical, Normal Academic and Express stream. The cluster with the highest proportion of high performers (Express stream) was labelled as "high performing". Input variables are then compared across the clusters to define the dominating traits of students in each performance group.

For continuous variables like PCI, Analysis of Variance (ANOVA) was used for mean comparison and Kolmogorov-Smirnov test (KS test) for overall distribution comparison. For nominal variables, Chi-squared test and Exact Binomial test (for binary variable) were used to detect significant differences. For ordinal variables, rank correlation test was conducted using Goodman and Kruskal's gamma.

Findings

Significance of the Variables

Measuring variable importance in an unsupervised model like cluster analysis requires working backward, by setting the cluster assignment as the target and then construct supervised models like random forest to predict that target. Mean Decreased Gini (MDG) can then be used to quantify the importance of input variables in contributing to an accurate prediction of the target, but also help to estimate how these variables play their part in driving the formation of clusters.

Table 1 lists the top significant input variables. Socioeconomic variables dominated the list. The most is the PCI followed by Fee Wavier (a proxy of household income). Household and housing variables is the next important group, further segregating the group of students by different family composition and living environment.
With this pilot group, and generalising to the larger population of MTS students, the key variables are mainly from the differentiation in socioeconomic status of these individuals.

Table 1. Mean Decrease Gini (MDG) and relative importance of clustering variables

Variables	MDG	Relative Importance
PCI	124.76	1.000
Fee waiver	117.68	0.943
Residence	50.31	0.403
No. of household members	25.32	0.203
No. of other children in household enrolled in MENDAKI	18.63	0.149
Rental status	4.27	0.034
Marital status (of main applicant)	3.18	0.025
Programme	2.91	0.023
Family role of main applicant	2.53	0.020
No. of household members aged 6 and below	2.49	0.020

Clusters Profile

Table 2. Characteristics of each cluster

	Cluster size	***Input Variables***	***PSLE Aggregate***
Cluster 1	273 (31.1%)	- Higher proportion having married main applicant - Higher proportion having smaller household sizes - Higher proportion with lesser children in MTS - Higher income distribution and mean - Higher proportion without fee waiver - Higher proportion living in non-rental block - Higher proportion living in larger houses	Higher proportion of high performers
Cluster 2	602 (68.8%)	- Higher proportion having unmarried main applicant - Higher proportion having larger household sizes - Higher proportion with more children in MTS - Lower income distribution and mean - Higher proportion having fee waiver - Higher proportion living in rental block - Higher proportion living in smaller houses	Higher proportion of low performers

Cluster 1 comprised of ~47% of students who are eligible for the Express stream, ~8% eligible for Normal Technical stream, while Cluster 2 has ~41% of students eligible for the Express stream and ~15% eligible for Normal Technical stream. Comparing between the two clusters, the differences are statistically significant, with Cluster 1 has higher proportion of high performers while Cluster 2 has higher proportion of low performers.
Statistical tests were conducted to detect significant differences between the two clusters. In terms of family structure. Some key findings include

- Cluster 1 has significantly higher proportion of students having married main applicant (~93%), as the main applicant are most often the parents, this functions proxy for intact family.
- ~42% of Cluster 1 are households with one to three members, while for Cluster 2, ~70% of them have at least 5 household members.
- Cluster 1 has a mean PCI of S$993.78, and less than 1% of the students have fee waiver as well as Education Trust Fund (ETF) (not used as input but has a correlation coefficient of almost 1 with fee waiver). On another hand, Cluster 2 has a mean PCI of S$296.82, demonstrating greater financial challenges as ~86% of the students were receiving financial aid.
- For residence status, all students living in rental block are clustered in Cluster 2, which also has ~85% staying in 4-room houses or below. While ~66% of Cluster 1 lives in 5-room houses or larger.

The results indicated a strong association between MTS students' academic performance in school and their socioeconomic background. The inability to perform well academically is the outcome of complicated interaction between many happenings in the student's life, and not solely because he/she is academically strong. Coming from a family with stable financial status means better access to learning resources, more conducive learning environment at home, parents have more time to accompany their children and many more.

Conclusion

Completing the MENDAKI pilot study is not a guarantee of the success of the full-scale study. However the pilot provided a better understanding of the phenomenon. The pilot has shown that the set of SES variables are key moderators of education attainment among Malay Muslims in Singapore. The pilot findings are also limited by the range of variables, there is a need to expand the variables to include other proxies of teaching and learning. The SES variables interfere with various factors that correlated with optimising educational performance. Future analysis should explore the impacts of socio-economic class on individual cultural capital, values and norms, conduciveness of home environment, time use, and access to external educational resources. It is quite clear that the problem of the under attainment of the Malay community cannot be solved by targeting only the students. This is indeed a difficult issue to tackle. It is inevitable that social and income inequalities will seep into the educational arena, resulting in a divergence of educational performance between more privileged students and less privileged students. It is important to understand the circumstance behind this phenomenon and the effects it has on educational attainment. While MENDAKI develops and enhances the MTS, there is a need to encompass a wider range student support that consisted of various types of educational, social and workforce development. Based on the pilot study's findings, the larger study adopted a sequential explanatory approach with a data mining phase where predictive analytics is applied to identify factors that influence as well as predict the academic performance of students in the MTS. This is followed by a qualitative case study approach with interviews and observations to better understand the impact of the SES variables.

References

Asiaone News (2007). *PM Lee's speech at Mendaki's 25th anniversary dinner and awards presentation. (2007, September 2).*

Channel NewsAsia. (2018, 14 May). *Malay-Muslim community must tackle 3 'elements' to ensure continued progress, says Masagos Zulkifli.* Accessed 10 September 2019, from https://www.channelnewsasia.com/news/singapore/malay-muslim-community-tackle-3-elements-progress-masagos-10232318

Department of Statistics, Singapore (2015) *Yearbook of Statistics.* Singapore: Department of Statistics.

Ministry of Education (MOE). (2018). *Education Statistics Digest 2018*. Accessed 10 September 2019, from https://www.moe.gov.sg/docs/default-source/document/publications/education-statistics-digest/esd_2018.pdf.

Tan, J., & Ho, B. T. (2001) 'A' levels or a polytechnic diploma? Malay students' choices of post-secondary options, in J. Tan, S. Gopinathan & W. K. Ho (Eds.), *Challenges Facing the Singapore Education System Today* (Singapore: Prentice Hall), 207-223.

Toh, Y.C. (2017, 1 April). Malay-Muslim groups can play key role to uplift the community. *The Straits Times*. Accessed 10 September 2019, from https://www.straitstimes.com/singapore/malay-muslim-groups-can-play-key-role-to-uplift-the-community

Yayasan MENDAKI (2017). About MENDAKI. Accessed 10 September 2019, from http://www.mendaki.org.sg/about-mendaki/general

Zhang, JY. (2014). Debunking the Myth of the Lazy Malays. *Mendaki Occasional Paper Series 1*. Accessed 10 September 2019 from http://www.mendaki.org.sg/qws/slot/u50178/2.%20Publication%20&%20Resources/2.5%20Others/2.5.4%20YM%20Occasional%20Papers/MENDAKI%20Occasional%20Paper%20Series_Zhang%20Jiayi.pdf

Please cite as: Mustapha, S. A., Li, J. Z. & Chong, S. (2019). Employing data to enhance teaching and learning in MENDAKI Tuition Scheme (MTS).In Y. W. Chew, K. M. Chan, and A. Alphonso (Eds.), *Personalised Learning. Diverse Goals. One Heart. ASCILITE 2019 Singapore* (pp. 354-358).