SUPPORT DOCUMENT

NCVER

What factors explain the likelihood of completing a VET qualification?

Adrian Ong and Michelle Circelli
National Centre for Vocational Education Research

INFORMING + INFLUENCING
THE AUSTRALIAN VET SECTOR

Publisher's note

The views and opinions expressed in this document are those of NCVER and do not necessarily reflect the views of the Australian Government, or state and territory governments. Any interpretation of data is the responsibility of the author/project team.

To find other material of interest, search VOCEDplus (the UNESCO/NCVER international database <http://www.voced.edu.au>) using the following keywords: completion; learning experience; outcomes; participation; qualifications; students; vocational education and training

This document should be attributed as Ong, A & Circelli, M 2018, *What factors explain the likelihood of completing a VET qualification?* NCVER, Adelaide.

This work has been produced by NCVER on behalf of the Australian Government and state and territory governments, with funding provided through the Australian Government Department of Education and Training.

COVER IMAGE: GETTY IMAGES/iStock

ISBN 978-1-925717-17-4
TD/TNC 131.06

Published by NCVER, ABN 87 007 967 311

Level 5, 60 Light Square, Adelaide, SA 5000
PO Box 8288 Station Arcade, Adelaide SA 5000, Australia

Phone +61 8 8230 8400 **Email** ncver@ncver.edu.au
Web <https://www.ncver.edu.au> <https://www.lsay.edu.au>
Follow us: <https://twitter.com/ncver> <https://www.linkedin.com/company/ncver>

Contents

Tables and figures

Tables

Figures

Executive summary

People participate in vocational education and training (VET) for a variety of reasons and at different stages of their life. Some undertake VET to gain the vocational skills necessary to enter the labour market for the first time, while others enter in order to upgrade existing skills, learn new ones, or simply for personal interest.

Successful completion of a VET qualification may not be the prime objective for all students. This consideration, together with the fact that not all people are equally capable of coping with the education and training demands required of some qualifications, suggests that measures of VET qualification completion rates may not be adequate for determining the full effectiveness of the sector. Hence, a number of different performance measures exist. However, little information is available on the likelihood of success for individual students or on the characteristics of those students more or less likely to succeed in completing their qualification. Consequently, there is a need to identify the various learner groups undertaking VET and determine those factors that impact upon their likelihood of success in completing their qualification.

Complementary to the publication *Australian vocational education and training statistics: VET program completion rates 2011—15,* the aim of this project is to identify the factors affecting the likelihood of completing a VET qualification among government-funded students. In doing so it is hoped that the findings prompt discussion on ways to improve VET completion by identifying the characteristics of those students most likely to complete a VET qualification. A further aim of this research is to explore the feasibility of using advanced data analytics to examine the factors that influence the likelihood of completing a VET qualification.

Method

To identify the important factors in explaining VET qualification completion, we used Classification and Regression Tree (CART) analysis, a form of decision tree learning.

Results

This analysis revealed that the top 10 factors[1] that explain the likelihood of completing a VET qualification are:

- course field of education
- labour force status
- course qualification level
- mode of attendance
- client apprenticeship flag (whether the course was part of an apprenticeship or traineeship)
- training provider type
- whether the course was commenced full-time

1 Every factor considered in the analysis was based on the last known enrolment activity, with the exception of age and whether the course was commenced full-time, which were based on the time of course commencement.

- training package flag
- state/territory that administered the funding of the training activity
- reason for undertaking the training

Our research also reveals that (in no particular order):

- Disadvantaged students (that is, Indigenous students, students with a disability and students from a low socioeconomic [SES] background) have a lower likelihood of completion.
- Students less likely to complete tend to be those enrolled in a certificate I or II qualification.
- Conversely, students in an apprenticeship or traineeship or who enrol full-time are more likely to complete the VET qualifications.
- Additionally, the use of multiple modes of learning increases the likelihood of completion.

Introduction

In this project, we examine the factors affecting the likelihood of completing vocational education and training qualifications[2]. It is hoped that the findings prompt discussion on ways to improve VET completion by identifying those students most likely to complete.

The focus of this research is on government-funded students who commenced their courses in 2011 or 2012. A total of 2.4 million course enrolment records, sourced from the National VET Provider Collection, are available for these two years, with almost all of the course enrolments part of nationally recognised VET courses. All course enrolments were at certificate I level and above.

Scope of analysis

Our data consists of government-funded students who commenced their courses in 2011 or 2012. The definition used is the same as that used in the *Australian vocational education and training statistics: government-funded students and courses 2016.*[3]

Choosing 2011—12 to analyse for completions was a simple decision to make (working backwards from 2016) given that it is reasonable to assume it would typically take a student four to six years to complete a VET qualification. Hence, within our population frame, two categories of students exist:

- completers: students who commenced a qualification in either 2011 or 2012 and who were subsequently awarded the qualification between 2011 and 2016
- not yet completers: students who commenced a qualification in either 2011 or 2012 but whom we have no information on their completion status between 2011 and 2016.

Project limitations

In the process of this project, we identify three primary challenges:

- The National VET Provider Collection does not collect information about course duration. Hence, length of study is not factored into the data analysis.
- A student's progression and articulation to a higher level of education (for example, from a foundation course such as certificate I to a certificate III) is not included in the analysis. For the purpose of the data analysis and reporting, only unique VET qualification enrolments are considered. Therefore, cases of students articulating to a higher qualification are viewed as separate course enrolments.
- Each state and territory is unique and each has certain artefacts that may not be captured in our data analysis. For example, some jurisdictions have certain reporting requirements, which make comparisons across states and territories difficult.

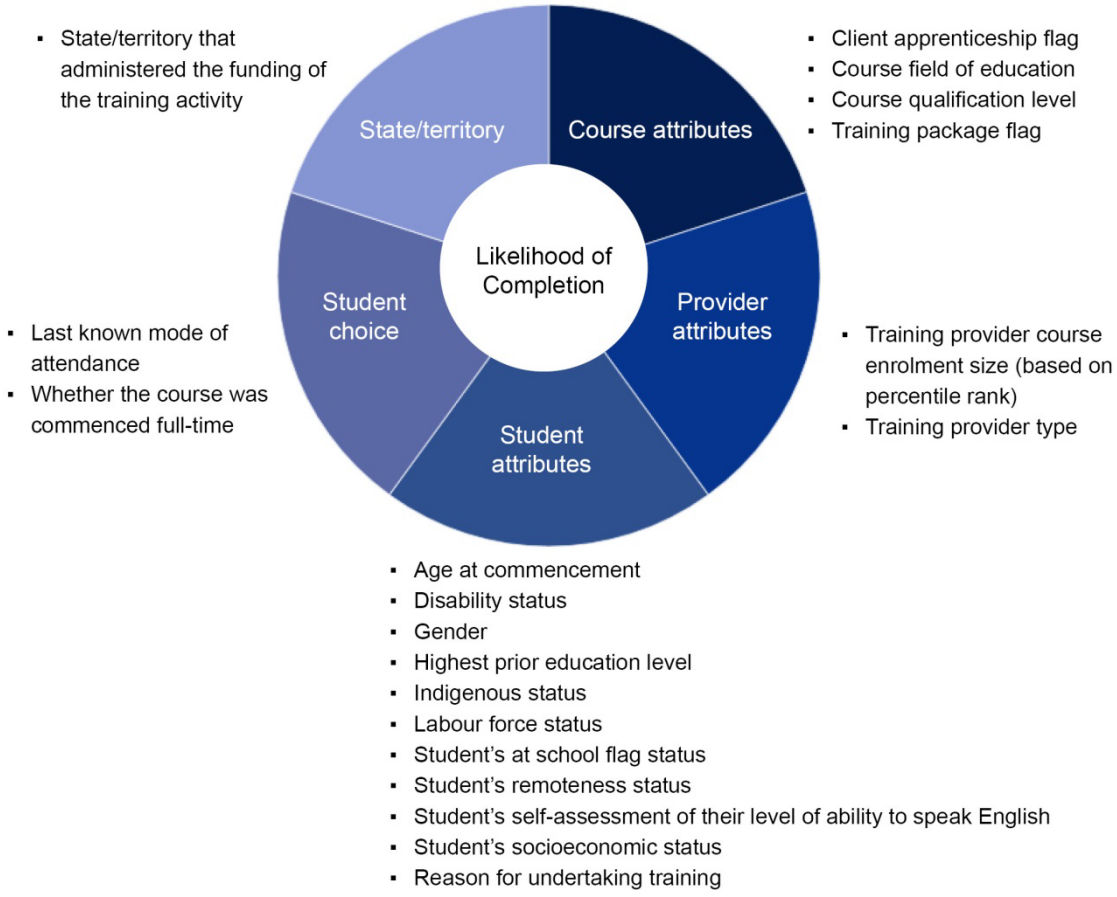
2 We use the term 'course' and 'qualification' interchangeably in this document.

3 Government-funded VET activity is defined as all Commonwealth and state/territory government-funded training delivered by technical and further education (TAFE) institutes, other government providers (such as universities), community education providers and other registered providers (such as privately operated registered training providers, schools, industry associations and enterprise providers). All fee-for-service activity from training providers has been excluded from the analyses reported here.

Factors affecting VET completion: a framework

The findings of previous research (John 2004) were confirmed by our exploration and analysis of the data. These indicate that the individual predictors for the likelihood of completion can be best summarised according to five overarching factors, which we refer to as the VET completion ecosystem (figure 1).

Figure 1 The VET completion ecosystem: the overarching factors[4] affecting the likelihood of completion



The individual predictors[4] examined within each factor are:

- student choice
  - last known mode of attendance: classroom-based only, electronic-based only, employment-based only, others (for example, correspondence), recognition of prior learning only, multiple modes of learning
  - whether the course was commenced full-time

4 Every predictor variable used in the analysis was based on the last known enrolment activity, with the exception of age and whether the course was commenced full-time, which were based on the time of course commencement.

- provider attributes
  - training course enrolment size (we use the course enrolment size as a proxy and classify the training provider based on percentile ranking, where the lowest percentile refers to the largest course enrolments)
  - training provider type (TAFE [technical and further education] institutes, universities, community education providers, and other registered providers)
- course attributes
  - client apprenticeship flag (whether the course was part of an apprenticeship or traineeship)
  - course field of education
  - course level of education
  - training package flag (whether the course was part of a training package)
- state/territory (that is, the state/territory that administered the funding of the training activity)
- student attributes
  - age at commencement
  - disability status
  - gender
  - highest prior education level
  - Indigenous status
  - labour force status
  - student is at school flag
  - student's remoteness status
  - student's self-assessment of their level of ability to speak English
  - student's socioeconomic status
  - reason for undertaking the training

The classification code frame used for each of the independent variables is available at appendix A.

Methodology and key findings

A generalised logistic mixed regression was used to gain insights into the profile of students who were likely to complete a government-funded VET qualification. A further analysis, employing the Classification and Regression Tree (CART) technique, was then used to determine which factors are important in predicting course completion.

Generalised logistic mixed regression

The rationale for the use of this approach is that information in our data source is hierarchical in nature; that is, students are nested in the training providers. We have reason to believe that qualification completion is not only influenced by student-level characteristics, but also by the characteristics of where they studied and what they studied. If we decided to ignore the hierarchical nature of the data and treat these students as though they were independent, we would run the risk of obtaining unreliable statistics, those where the standard errors are under-estimated and the test statistics are over-estimated.

Table 1 gives the covariance parameter estimates from the regression analysis. As the estimate of the variance of the random intercept is significantly greater than 0, we can justify the use of this statistical method. Otherwise, an ordinary logistic regression model would have been sufficient.

Table 1 Covariance parameter estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	Provider Identifier, Course Identifier	2.65	0.045	59.2	<.0001

The Fit Statistics table, as shown in table 2, confirms that the variability in the data has been satisfactorily modelled as the ratio of the generalised chi-square statistic and its degree of freedom is close to 1. This indicates there is no substantial residual over-dispersion, which otherwise suggests bias in standard errors.

Table 2 Fit statistics from the regression model

Fit Statistics for conditional distribution	
-2 log L(COMPLETION_STATUS | r. effects)	2123305
Pearson Chi-Square	2121356
Pearson Chi-Square / DF	0.90

Likelihood of course completion

The overall predicted likelihood of course completion based on the 2011 and 2012 cohorts is 40.1%. The model indicates there is no significant difference between the 2011 and 2012 government-funded cohorts at the national level. A comparison with the published actual completion rate in the *Australian vocational education and training statistics: VET program completion rates, 2011—15* indicate that the disparity between the predicted and actual completion rates is negligible. Figure 2 shows these statistics.

Figure 2 Likelihood of government-funded VET qualification completion at national level

Note: 1 NCVER 2017, *Australian vocational education and training statistics: the likelihood of completing a government-funded VET program, 2011–15*, NCVER, Adelaide.
Source: NCVER 2016 National VET Provider Collection

Table 3 presents the results from the generalised logistic mixed regression model for the 2011 and 2012 cohorts. This table shows the predicted likelihood of completion as least squares means. These least squares means are the marginal mean for each factor, adjusted for other variables in the model.

Table 3 Predicted likelihood of government-funded VET qualification completion for the student cohort 2011 and 2012 (based on generalised logistic mixed regression)

Factors/attributes[5]		Level	Likelihood of completion (%) (Least squares mean)	Standard error of the mean (%)
Overall		Overall	40.1	1.6
Cohort		2011	40.2	1.7
		2012	39.9	1.6
Student choice	Last known mode of attendance	Classroom only	33.5	1.6
		Electronic only	29.9	1.6
		Employment-based only	36.1	1.9
		Other (e.g. correspondence)	30.7	1.6
		RPL/ credit transfer only	56.4	2.2
		Multiple modes	56.1	1.8

5 Unless otherwise stated, every factor considered in the analysis was based on the last known enrolment activity.

Table 3 Cont.

Factors/attributes[5]		Level	Likelihood of completion (%) (Least squares mean)	Standard error of the mean (%)
	Whether the course was commenced full-time	No	26.3	1.3
		Yes	55.6	1.7
Provider attributes	Provider type	Technical and further education institutes	41.7	1.6
		Universities	32.2	2.0
		Community education providers	42.7	2.0
		Other registered providers	44.1	1.6
Course attributes	Client apprenticeship flag	Not part of an apprenticeship or traineeship	30.2	1.4
		Part of an apprenticeship or traineeship	50.8	1.9
	Course field of education	Natural and physical sciences	41.2	3.4
		Information technology	42.2	2.7
		Engineering and related technologies	38.7	1.9
		Architecture and building	40.6	3.3
		Agriculture, environmental and related studies	35.2	1.8
		Health	49.7	2.1
		Education	46.0	3.1
		Management and commerce	45.6	1.8
		Society and culture	51.4	2.0
		Creative arts	33.8	2.1
		Food, hospitality and personal services	37.7	1.9
		Mixed field programs	22.3	1.8
	Course qualification level	Diploma and above	43.9	1.8
		Certificate IV	41.9	1.7
		Certificate III	45.0	1.7
		Certificate II	39.1	1.8
		Certificate I	31.0	2.2
State/territory	State/territory that administered the funding of the training activity	New South Wales	35.9	1.9
		Victoria	40.4	1.7
		Queensland	39.6	2.1

Table 3 Cont.

Factors/attributes[5]		Level	Likelihood of completion (%) (Least squares mean)	Standard error of the mean (%)
		South Australia	43.5	2.5
		Western Australia	32.8	1.7
		Tasmania	49.6	2.9
		Northern Territory	46.9	3.7
		Australian Capital Territory	32.8	3.3
Student attributes	Gender	Female	46.2	1.4
		Male	41.5	1.4
	Age at commencement	Under 19 years old	37.0	1.6
		20–29 years old	38.8	1.6
		30–39 years old	41.9	1.7
		40–49 years old	43.8	1.7
		50 years and above	43.4	1.7
	Disability status	Student with a disability	35.7	2.4
		Student without a disability	42.3	1.7
	Indigenous status	Indigenous	33.4	1.6
		Non-Indigenous	44.6	1.7
	Socioeconomic status	Low socioeconomic status	38.5	1.8
		Medium socioeconomic status	39.6	1.8
		High socioeconomic status	41.1	1.8
	Student's remoteness status	City	41.2	1.6
		Regional	41.7	1.6
		Remote	39.1	1.6
		Overseas	49.0	4.0
	At school status	At school flag = No	38.7	1.6
		At school flag = Yes	40.7	1.7
	Prior education at the time of course commencement	Did not have prior education at the time of course commencement	37.8	1.6
		Had prior education at the time of course commencement	42.3	1.7

Table 3 Cont.

Factors/attributes[5]		Level	Likelihood of completion (%) (Least squares mean)	Standard error of the mean (%)
	Labour force status	Full-time employee	43.7	1.7
		Part-time employee	42.4	1.7
		Self-employed – not employing others	43.0	1.7
		Employer	39.9	1.8
		Employed – unpaid worker in a family business	40.0	1.7
		Unemployed – seeking full-time work	38.4	1.6
		Unemployed – seeking part-time work	38.6	1.6
		Not employed – not seeking employment	38.7	1.6
	Study reason	Employment-related reason	41.2	1.7
		Further study reason	40.2	1.8
		Personal and other reason	38.7	1.6

Source: NCVER 2016 National VET Provider Collection

The inclusion of both the 2011 and 2012 cohorts in the data analysis enables a test to determine whether there is any significant difference between the two cohorts. The following table 4 summarises the key findings, which indicates a significant difference between students who are more likely to complete a VET course as opposed to those who, four to five years after commencing their training, are yet to complete.

Table 4 Key findings on students from the 2011 and 2012 cohort on their likelihood of completing a government-funded VET qualification

Student choice	Provider attributes
▪ Full-time study (56%) versus non full-time study (26%) ▪ Mode of attendance (last known): multiple modes of learning (56%) versus those who were enrolled purely in classroom (34%), electronic-based learning (30%), employment-based learning (36%) or correspondence learning (31%)	▪ Community education providers (43%) versus universities (32%) ▪ TAFE (42%) versus universities (32%) ▪ Other RTOs (44%) versus TAFE (42%)
Course attributes	**Student attributes**
▪ Being an apprentice or trainee (51%) versus not being an apprentice or trainee (30%) ▪ Diploma and above (44%) versus those enrolled in certificate I (31%), certificate II (39%) and certificate IV (42%) ▪ Certificate III (45%) versus certificate IV (42%) ▪ Students enrolled in mixed field courses versus other fields of education	▪ Non-Indigenous (45%) versus Indigenous (33%) ▪ Females (46%) versus males (41%) ▪ Students with prior education at the time of course commencement (42%) versus students without any prior education (38%) ▪ Employed full-time (44%) versus unemployed — seeking full-time work (38%), unemployed — seeking part-time work (39%) and not employed (not seeking employment) (39%)

Source: NCVER 2016 National VET Provider Collection

Decision tree technique (Classification and Regression Tree)

The Classification and Regression Tree (CART) technique[6], a decision tree learning technique, was used to compute the relative decisive power (importance) of each factor and how much it is influencing the likelihood of completing a government-funded VET qualification.

Key advantages of the CART technique are:

- Its main strength lies with the fact that the output shows the factors that are important to the model in terms of their explanatory power and variance.
- CART is iterative and it re-evaluates other variables continuously to build the tree and hence allows non-linear relationships between the variables.
- CART is easy to set up to train the data (i.e. learn and discover potential predictive relationships) and cross validate the model building[7].
- CART can be adjusted for the level of data misclassification during the model development.
- It produces a decision tree diagram that shows the various student segments/clusters that are important in predicting the likelihood of course completion.

Core to the CART algorithm is the concept of 'purity' and 'impurity' in each leaf node[8]. In particular, a Gini coefficient (a measure of dispersion) is used as a splitting criterion in constructing the decision tree. The Gini aims to maximise the homogeneity of the leaf nodes with respect to the targeted outcome variable, hence making substantive reduction in 'impurity' as its goal in building the decision tree.

The CART algorithm was run on the 2011 and 2012 cohorts separately. However, as this is a preliminary and exploratory piece of work, we only consider and present the 2011 cohort in this paper. While there are arguments to include the 2012 cohort, we feel that to give it appropriate consideration, the same number of years of activity should be included in the analysis. Furthermore, there may be inherent differences between the two cohorts that may not yet be apparent.

In implementing the CART algorithm on the 2011 cohort, we set the 'misclassification cost' higher for students who completed the qualification. This approach was intentional because we know for certain that a student has graduated as this information is captured in the National VET Provider Collection. Conversely, the National VET Provider Collection has no information about whether a student has dropped out of the training.

The overall misclassification risk is about 30% with the predicted classification accuracy of 69%.

Table 5 Misclassification risk for the 2011 data

Population frame	Method	Mean	Standard error
2011	Resubstitution	0.296	<0.001
2011	Cross validation	0.296	<0.001

Source: NCVER 2016 National VET Provider Collection

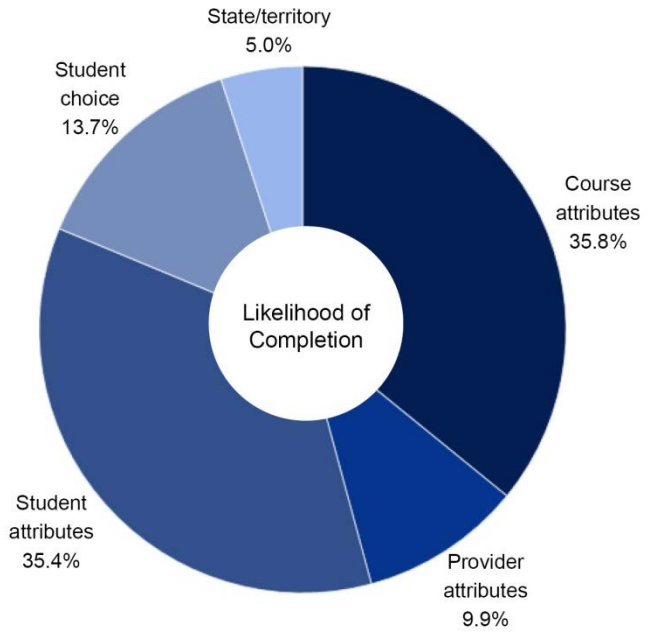
6 There is an extensive literature written on this technique, see for example Breiman et al. (1984), or Loh (2008).
7 A method used to assess the predictive models by dividing the original sample into a training dataset to train the model, and a test dataset to evaluate it.
8 Leaf nodes are the segments formed in the decision tree.

In table 6 we provide the computed overall relative contributing importance score for each factor. This table shows the normalised importance score, where the largest measure of importance has a score of 100. The magnitude of importance for the remaining factors is then compared against this 100-mark baseline.

The factors[9] in order of importance to the model for determining government-funded VET qualification completion for the 2011 cohort are:

1 course field of education
2 labour force status
3 course qualification level
4 mode of attendance
5 client apprenticeship flag (whether the course was part of an apprenticeship or traineeship)
6 training provider type
7 whether the course was commenced full-time
8 training package flag (whether the course was part of a training package)
9 state/territory that administered the funding of the training activity
10 reason for undertaking the training

Figure 3 Contributing factors to the likelihood of completing a government-funded VET qualification, 2011 cohort



Source: Government-funded students who commenced their VET qualification in 2011. NCVER 2016 National VET Provider Collection

9 Every factor considered in the analysis was based on the last known enrolment activity, with the exception of age and whether the course was commenced full-time, which were based on the time of course commencement.

Table 6 Factors[1] contributing to the likelihood of government-funded VET qualification completion, 2011 cohort

	Overall relative contributing importance score 2011 cohort	Normalised importance score 2011 cohort
Student choice		
Last known mode of attendance	0.020	58.0
Whether the course was commenced full-time	0.016	46.4
Provider attributes		
Training provider type	0.019	54.8
Training provider course enrolment size *(based on percentile rank)*	0.007	20.7
Course attributes		
Course field of education	0.035	100.0
Course qualification level	0.025	72.4
Client apprenticeship flag	0.020	57.4
Training package flag	0.015	42.5
State/territory		
State/territory that administered the funding of the training activity	0.013	38.3
Student attributes		
Labour force status	0.030	85.9
Reason for undertaking training	0.012	34.8
Student's at school flag status	0.011	30.6
Disability status	0.010	28.6
Age at commencement	0.008	23.8
Highest prior education level	0.007	19.7
Student's remoteness status	0.006	17.2
Indigenous status	0.005	15.0
Student's self-assessment of their level of ability to speak English	0.002	6.7
Student's socioeconomic status	0.001	3.8
Gender	0.001	3.2

Note: 1 Every factor considered in the analysis was based on the last known enrolment activity, with the exception of age and whether the course was commenced full-time, which were based on the time of course commencement.

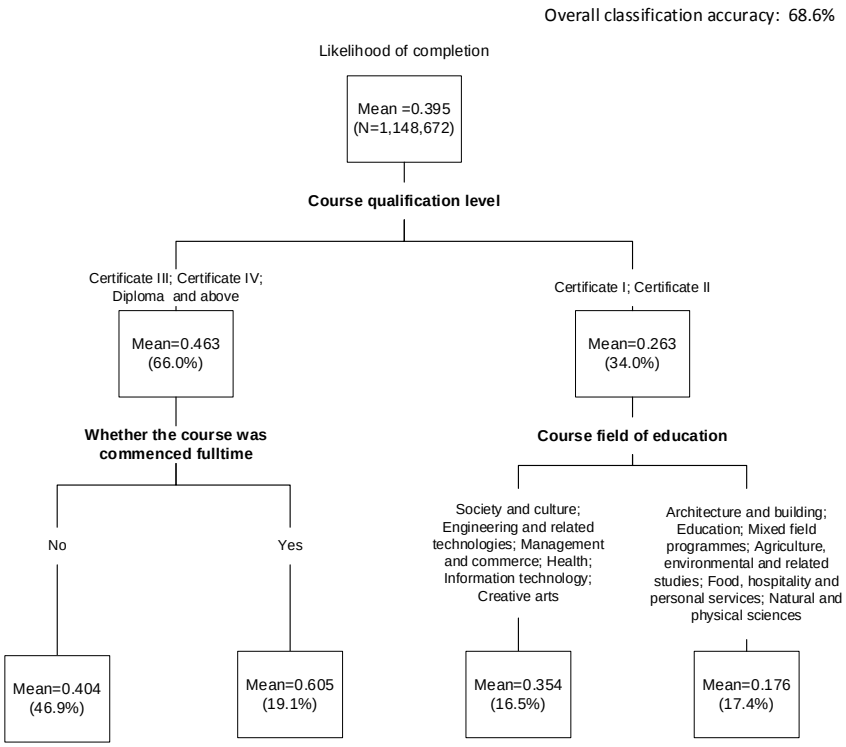
Source: Government-funded students who commenced their VET qualification in 2011. NCVER 2016 National VET Provider Collection

Taking the results from CART and applying our earlier concept of the VET Completion Ecosystem (figure 1), it became clear that both the course and student attributes play a pivotal role in explaining the likelihood of completion at the aggregate level (figure 3). The percentages shown in figure 3 for the 2011 cohort reflect the degree of influence each individual factor has on the likelihood of completion. Using the results from the variable importance analysis (table 6), this pie chart shows the proportion of contribution the various factors have in determining likelihood of completion.

The CART technique also produces a decision tree. Figure 4 shows the decision tree up to three levels for the 2011 cohort. For a further breakdown, refer to appendix B. Appendix C shows the decision tree diagrams by state and territory.

The main idea behind the CART decision tree is that we form a binary tree and we minimise the error for each leaf node of a tree. The final aim of a decision tree is to identify leaf nodes that produce substantive reduction in impurity (see earlier discussion). Simply put, the first node indicates the overall likelihood (probability) of completing a government-funded VET qualification. The splitting of the subsequent nodes depends on which predicative variable is able to produce a substantive reduction in impurity, conditional on the previously assigned node.

Figure 4 Decision tree diagram of the likelihood of government-funded VET qualification completion, 2011 cohort



Note: Mean refers the likelihood (i.e. probability) of completing a government-funded VET qualification. The percentage figure inside the parenthesis refers to the cluster size relative to the population frame in scope (i.e. N).
Source: NCVER 2016 National VET Provider Collection.

Figure 4 shows two important statistics. The mean refers to the average likelihood of completion, while the percentage figure inside the parenthesis refers to the cluster size relative to the population frame in scope. The population frame in scope is inside the parenthesis in the top of the tree (N=1,148,672).

Here, we observe that the overall average likelihood of completion is 0.395 for the 2011 cohort at the national level. As qualification level produces the most substantive reduction in impurity, it ranked first in the list of factors. The CART found that students enrolled in certificate III and above are more likely to complete their VET qualifications (46.3%) than those enrolled in certificates I and II (26.3%).

Conditional on those who were enrolled in certificate III and above, full-time study status becomes an important attribute, whereby we observe studying full-time is likely to increase the students' mean probability of completing their VET qualifications (60.5%). On the other hand, those who were not enrolled full-time had a lower likelihood of completion, at 40.4%.

Among those who were enrolled in certificates I and II, course field of education was a key attribute. In particular, students had a higher chance of completing if they were enrolled in the fields of education of: Society and culture; Engineering and related technologies; Management and commerce; Health; Information technology; and Creative arts (35.4%).

Conclusion

The undertaking of this research provided the opportunity to explore the feasibility of using advanced data analytics while examining the factors that influence the likelihood of completing a VET qualification. Our approach evolved such that two statistical techniques were used to answer the key questions relating to the likelihood of completing a VET qualification. The two techniques are the generalised logistic mixed regression and the Classification and Regression Tree technique (CART) of machine learning.

Neither statistical technique is superior to the other. Each has a distinct approach and each technique assists in answering our research questions from different perspectives. The generalised logistic mixed regression model was used to answer the question on the likelihood of completion. The CART technique extends this scope to determine the characteristics that are most important in predicting course completion. The CART approach also has the capacity to illuminate the interaction effects between student characteristics, provider characteristics and course characteristics, as depicted in the decision tree diagrams.

The analyses presented here are preliminary but it does highlight areas for further discussion with respect to which student or training attributes can be the target for interventions to help increase the likelihood of completion. Furthermore, we believe the analyses are sufficiently developed to raise interest among various stakeholders and researchers in the use of data science to answer VET questions, especially in the area of completion rates.

A possible extension of this research is to subsequently analyse different years to fine-tune our model to get a clearer sense of what factors are influencing the likelihood of VET qualification completion.

References

Breiman, L, Friedman, JH, Olshen, RA & Stone, CJ 1984, *Classification and regression trees,* Wadsworth International Group, Belmont.

John, D 2004, *Identifying the key factors affecting the chance of passing vocational education and training subjects*, NCVER, Adelaide, viewed 31 July 2017, <https://www.ncver.edu.au/publications/1460.html>.

Loh, WY 2008, 'Classification and regression tree methods', in *Encyclopedia of statistics in quality and reliability,* eds F Ruggeri, RS Kenett & FW Faltin, pp.315–323, viewed 31 Jul 2017, <https://www.stat.wisc.edu/~loh/treeprogs/guide/eqr.pdf>

NCVER (National Centre for Vocational Education Research) 2015, *A preliminary analysis of the outcomes of students assisted by VET FEE-HELP*, NCVER, Adelaide, viewed 31 July 2017, <https://www.ncver.edu.au/publications/2826.html>.

NCVER 2017, *Australian vocational education and training statistics: the likelihood of completing a government-funded VET program, 2011–15*, NCVER, Adelaide, viewed 31 July 2017,

<https://www.ncver.edu.au/publications/publications/all-publications/vet-program-completion-rates-2011-15>

NCVER 2016 National VET Provider Collection, viewed 31 July 2017,

<https://www.ncver.edu.au/data/collection/students-and-courses-collection/government-funded-students-and-courses>

Appendix A – Terms and definitions

Course qualification level

The course level of education is based on the Australian Qualifications Framework (AQF). It is a unified system of national qualifications in schools, vocational education and training (TAFE institutes and private providers) and the higher education sector (mainly universities). AQF levels are an indication of the relative complexity and/or depth of achievement and the autonomy required to demonstrate that achievement.

- Diploma and above
- Certificate IV
- Certificate III
- Certificate II
- Certificate I

Course field of education

It identifies the subject matter that is the ultimate aim of the skills and knowledge gained in a qualification, course or skill set.

- 01 - Natural and physical sciences
- 02 - Information technology
- 03 - Engineering and related technologies
- 04 - Architecture and building
- 05 - Agriculture, environmental and related studies
- 06 - Health
- 07 - Education
- 08 - Management and commerce
- 09 - Society and culture
- 10 - Creative arts
- 11 - Food, hospitality and personal services
- 12 - Mixed field programs

State/territory that administered the funding of the training activity

Uniquely identifies the funding state or territory for the qualification.

- 1 - New South Wales
- 2 - Victoria
- 3 - Queensland

- 4 - South Australia
- 5 - Western Australia
- 6 - Tasmania
- 7 - Northern Territory
- 8 - Australian Capital Territory

Client apprenticeship flag (based on last known enrolment activity)

A flag that indicates whether a student is undertaking some training under an Apprenticeship/Traineeship Training Contract.

- Y - Client has at least one enrolment that is associated with an apprenticeship/traineeship training contract.
- N - Client does not have any enrolments that are associated with an apprenticeship/traineeship training contract.

Training package flag (based on last known enrolment activity)

- 1 - Course is part of the training package
- 0 - Course is not part of the training package

Training organisation provider type

It identifies the type of institution or organisation providing training to the student as reported by the training organisation.

- Technical and further education institute (TAFE)
- Universities
- Community education providers
- Other registered training providers

Training provider enrolment size

It is based on course enrolment percentile ranking where the smallest percentile means that the training provider has an extremely large number of course enrolments.

Students at school flag status (based on last known enrolment activity)

At school flag indicates whether a student is currently attending secondary school.

- Y – Yes
- N – No
- @ - No information

Disability status (based on last known enrolment activity)

A flag that indicates whether students consider themselves to have a disability, impairment or long-term condition.

- Y – Yes
- N – No
- @ - No information

Indigenous status (based on last known enrolment activity)

It indicates a client who self-identifies as being of Aboriginal or Torres Strait Islander descent.

- 1 - Indigenous
- 2 - Non-Indigenous
- 999 – No information

Prior education at the time of course commencement

It indicates if the student had prior education at the time of course commencement

- Did not have prior education at the time of course commencement
- Had prior education at the time of course commencement

Highest prior education level (based on last known enrolment activity)

It indicates the highest level of education completed by a student

- 008 - Bachelor degree or above
- 02 - Did not go to school
- 09 - Year 9 or lower
- 10 - Year 10
- 11 - Year 11
- 12 - Year 12
- 410 - Advanced diploma/Associate degree
- 420 - Diploma
- 511 - Certificate IV
- 514 - Certificate III
- 521 - Certificate II
- 524 - Certificate I
- 990 - Miscellaneous education
- *** - Unknown

Student's remoteness status (based on last known enrolment activity)

It identifies the level of remoteness of a location in terms of the ease or difficulty people face in accessing services in non-metropolitan Australia. The classification is based on the Australian Standard Geographical Classification-Remoteness Area.

- 0 - Major cities
- 1 - Inner regional
- 2 - Outer regional
- 3 - Remote
- 4 - Very remote
- 8 - Overseas
- 9 - No usual address
- Unknown - unknown

Reason for undertaking training (based on last known enrolment activity)

This derived field is based on the reason the student is undertaking the subject enrolment.

- 1 - Employment related
- 2 - Further study related
- 3 - Personal and other reasons
- 99 - Not stated

Student's socioeconomic status (based on last known enrolment activity)

This field identifies the socio-economic status of a student based on the Index of Relative Socio-Economic Disadvantage (IRSD)[10] classification. It is a general socio-economic index that summarises a range of information about the economic and social conditions of students within an area.

- 1 - Quintile 1: Most disadvantaged (IRSD decile 1 & 2)
- 2 - Quintile 2 (IRSD decile 3 & 4)
- 3 - Quintile 3 (IRSD decile 5 & 6)
- 4 - Quintile 4 (IRSD decile 7 & 8)
- 5 - Quintile 5: Least disadvantaged (IRSD decile 9 & 10)
- @ - Unknown (IRSD decile N/A)

10 2033.0.55.001 - Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2011

Gender (based on last known enrolment activity)

- M - Male
- F – Female
- @ - No information

Age at commencement

If age is reported incorrectly or not stated, it is coded as 999.

Student's self-assessment of their level of ability to speak English (based on last known enrolment activity)

- 01 - English
- 02 - English well and other language
- 03 - English not well and other language
- @ - Unknown

Last known mode of attendance

It captures the manner in which a student is undertaking the course during his/her last known enrolment activity.

- 1 - Classroom-based only (if all the enrolled subjects are classroom based)
- 2 - Electronic-based only (if all the enrolled subjects are electronic based. For example, web-based resources, computer-based resources, online interactions both on and off campus include radio, television, videoconference, or audio-conference)
- 3 - Employment-based only (if all the enrolled subjects are employment based)
- 4 - Other (e.g. correspondence) only (if all the enrolled subjects are reported as others – for example correspondence)
- 5 - RPL/CT only (if all the enrolled subjects have received recognition of prior learning or credit transfer)
- 6 – Multimodal (if the enrolled subjects are a mix of the above)

Whether the course was commenced full-time

- Y – Yes
- N – No

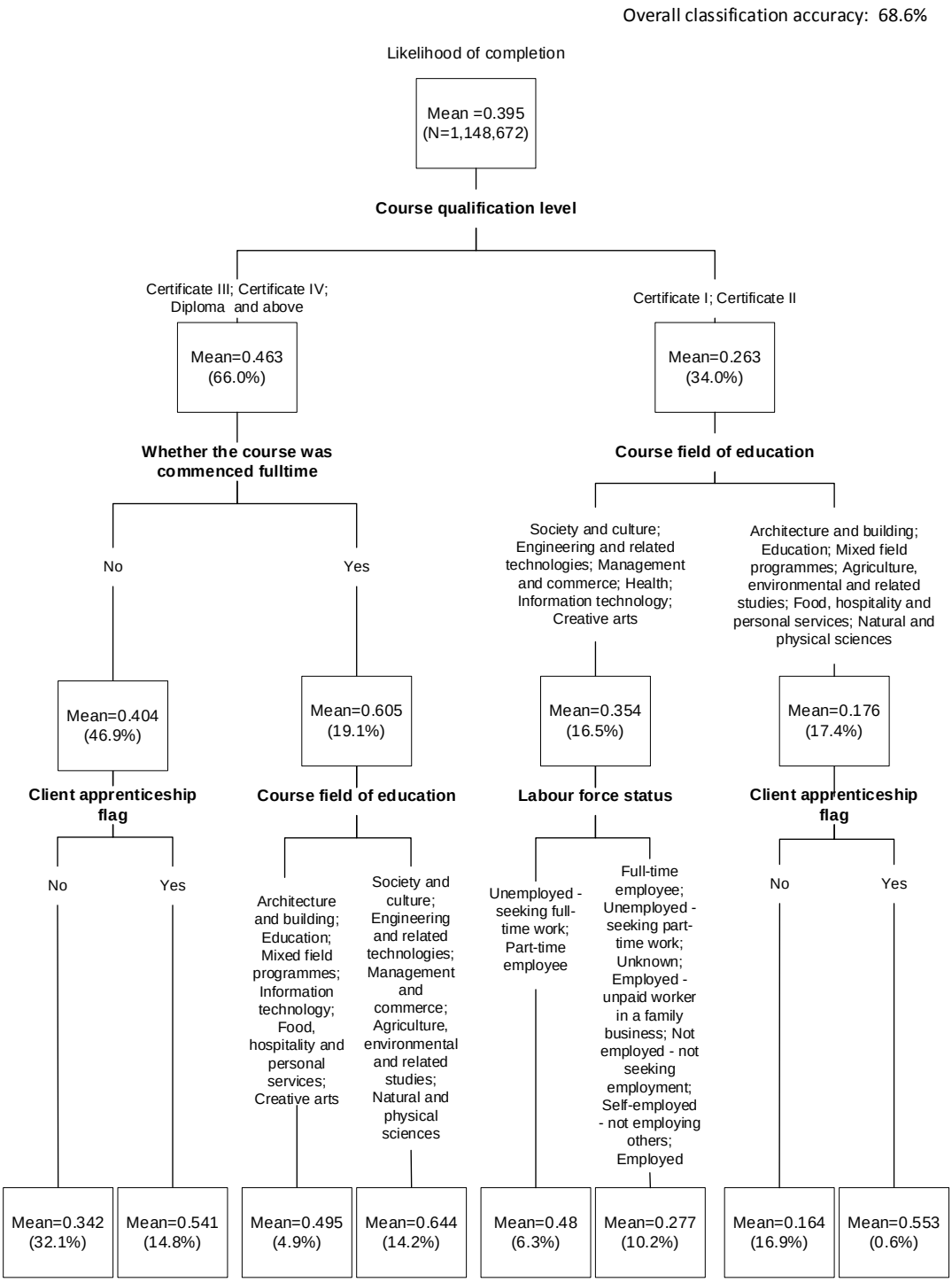
Labour force status (based on last known enrolment activity)

The labour force status identifier describes a student's employment status as captured on the student's enrolment form.

- 01 - Full-time employee
- 02 - Part-time employee
- 03 - Self-employed - not employing others
- 04 - Employer
- 05 - Employed - unpaid worker in a family business
- 06 - Unemployed - seeking full-time work
- 07 - Unemployed - seeking part-time work
- 08 - Not employed - not seeking employment
- @ - Unknown

Appendix B – Decision Tree Diagram (National)

Figure B1 Decision tree diagram on the likelihood (probability) of government-funded VET qualification completion, 2011 cohort

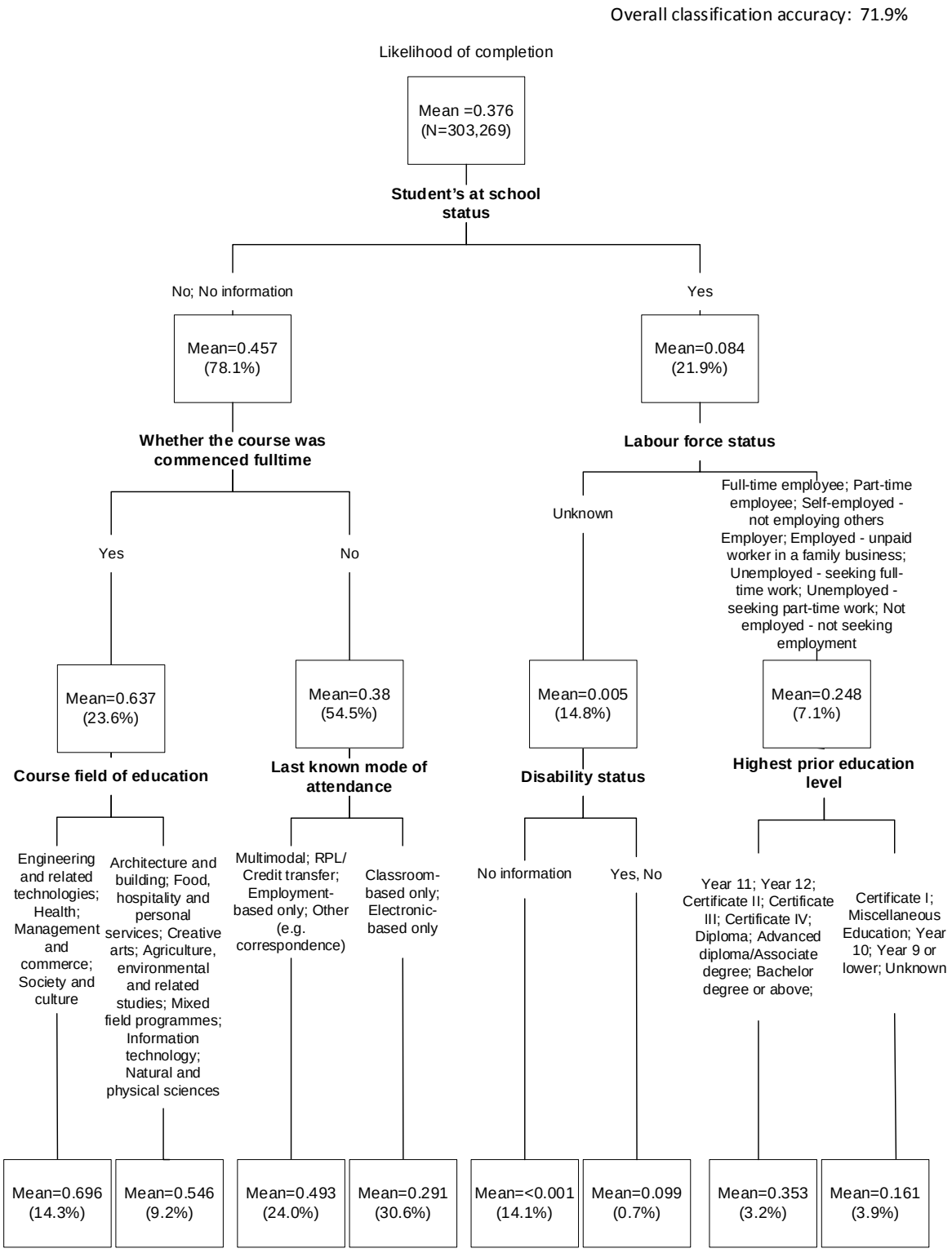


Note: Mean refers to the likelihood (i.e. probability) of completing a government-funded VET qualification. The percentage figure inside the parenthesis refers to the cluster size relative to the population frame in scope (i.e. N).

Appendix C – Decision tree diagrams (states and territories that administered the funding of the training activity)

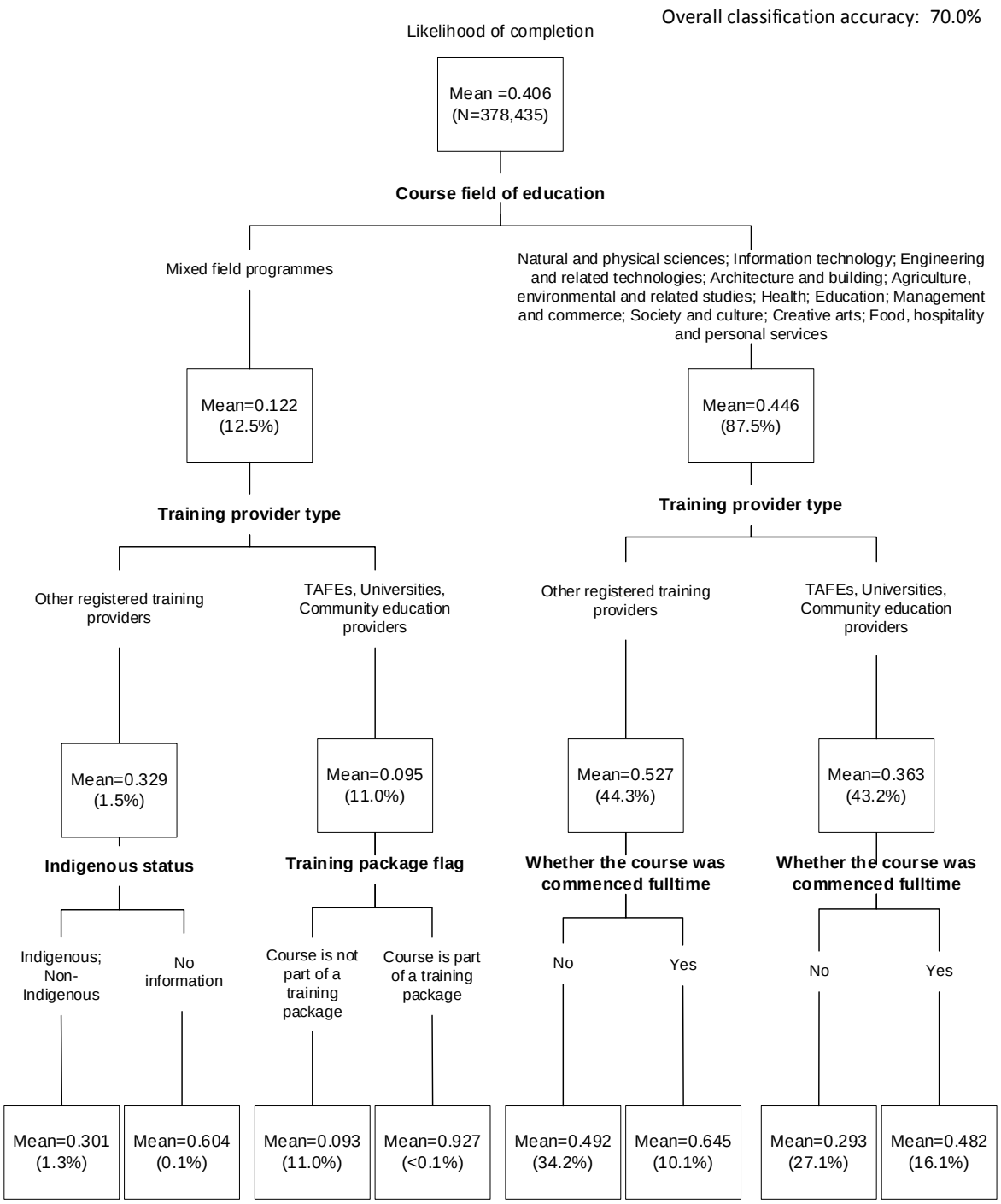
A decision tree is a flow chart that includes a root node, branches, and leaf nodes. The first node (root node) indicates the overall likelihood (probability) of completing a government-funded VET qualification. The splitting of the subsequent nodes depends on which predicative variable is able to produce a substantive reduction in impurity, conditional on the previously assigned node. Each leaf node shows the likelihood (probability) of course completion, and the student cluster size with respect to the population frame of the root node.

Figure C1 Decision tree diagram for New South Wales 2011 cohort



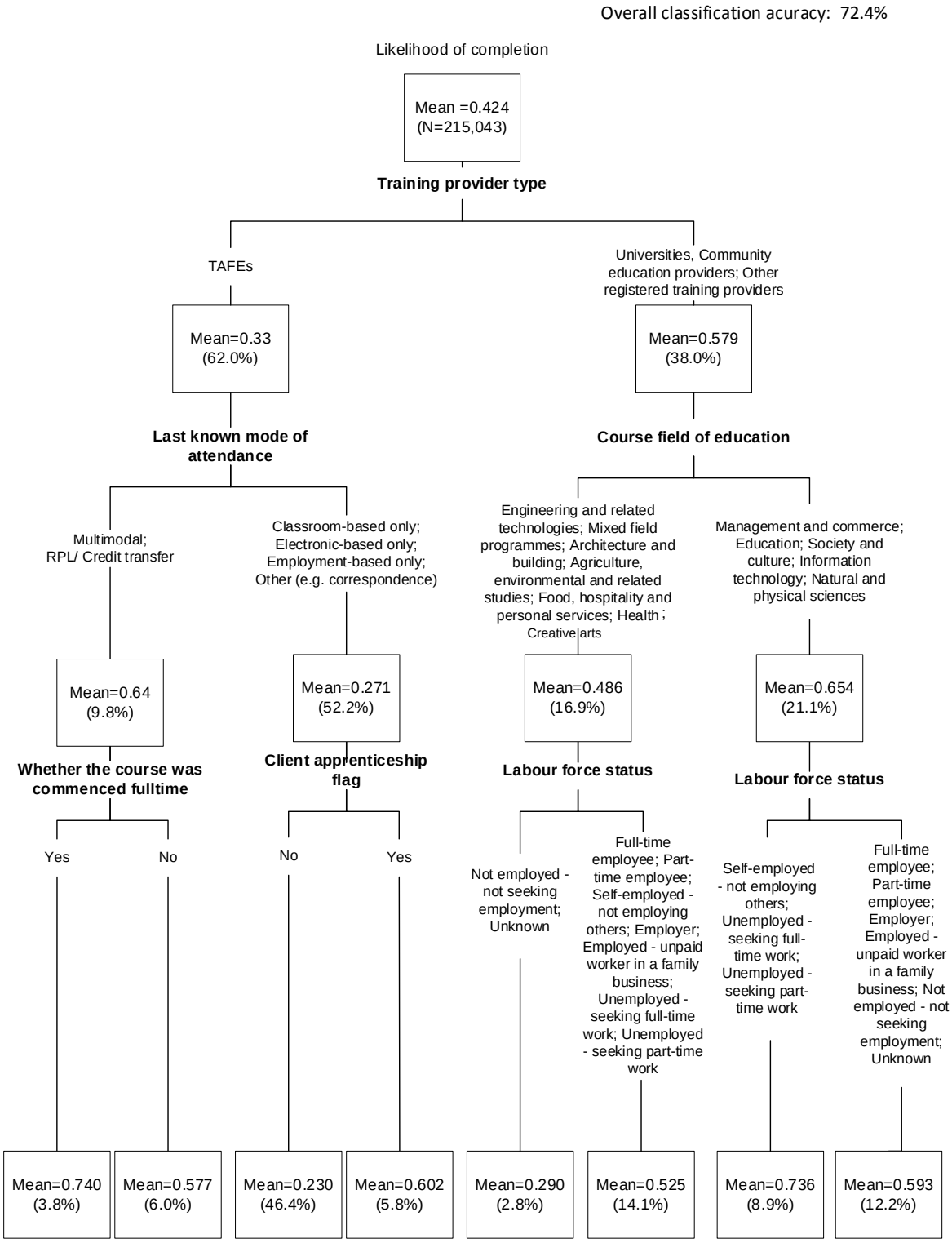
Note: Mean refers to the likelihood (i.e. probability) of completing a government-funded VET qualification. The percentage figure inside the parenthesis refers to the cluster size relative to the population frame in scope (i.e. N).

Figure C2 Decision tree diagram for Victoria 2011 cohort



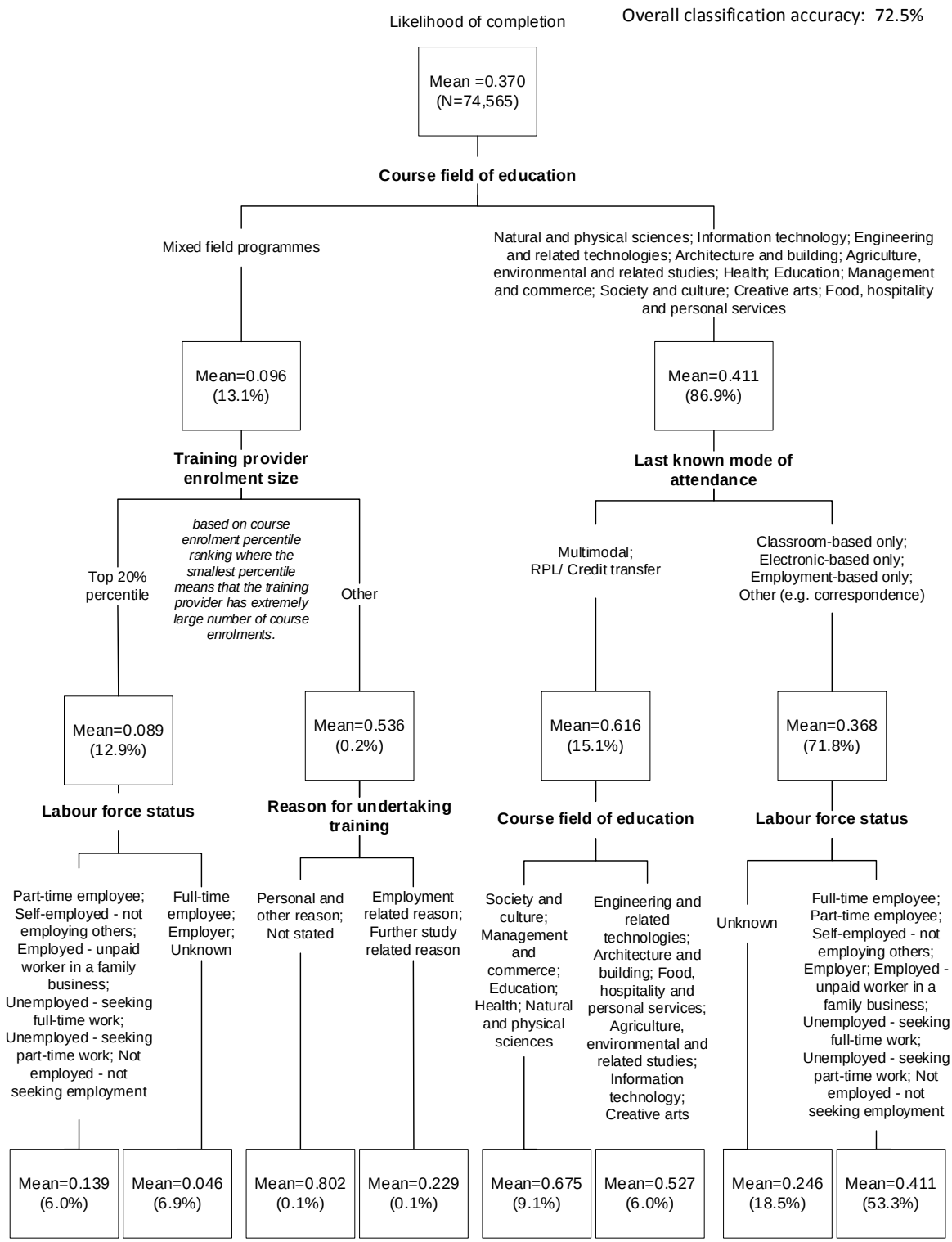
Note: Mean refers to the likelihood (i.e. probability) of completing a government-funded VET qualification. The percentage figure inside the parenthesis refers to the cluster size relative to the population frame in scope (i.e. N).

Figure C3 Decision tree diagram for Queensland 2011 cohort



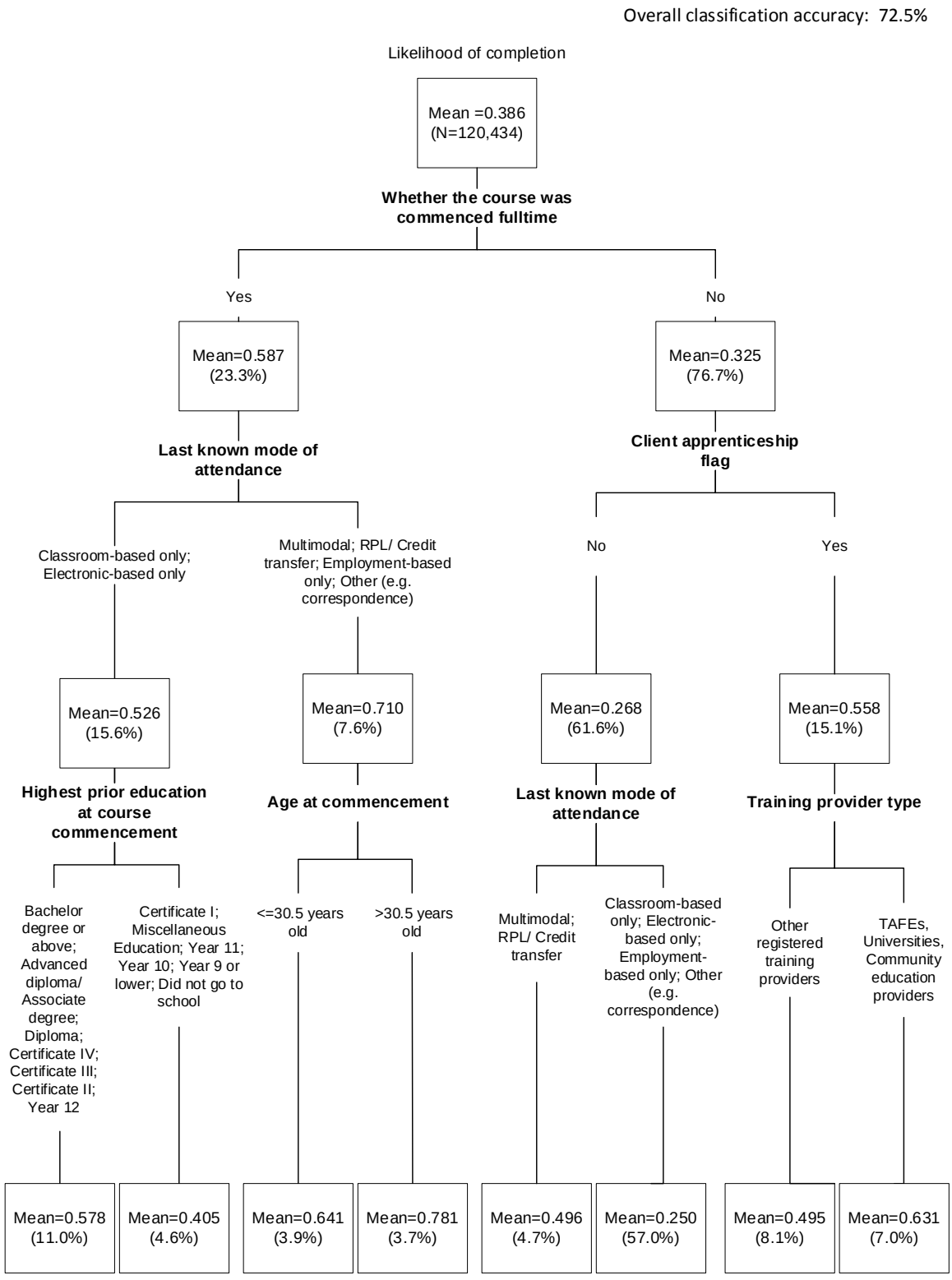
Note: Mean refers to the likelihood (i.e. probability) of completing a government-funded VET qualification. The percentage figure inside the parenthesis refers to the cluster size relative to the population frame in scope (i.e. N).

Figure C4 Decision tree diagram for South Australia 2011 cohort



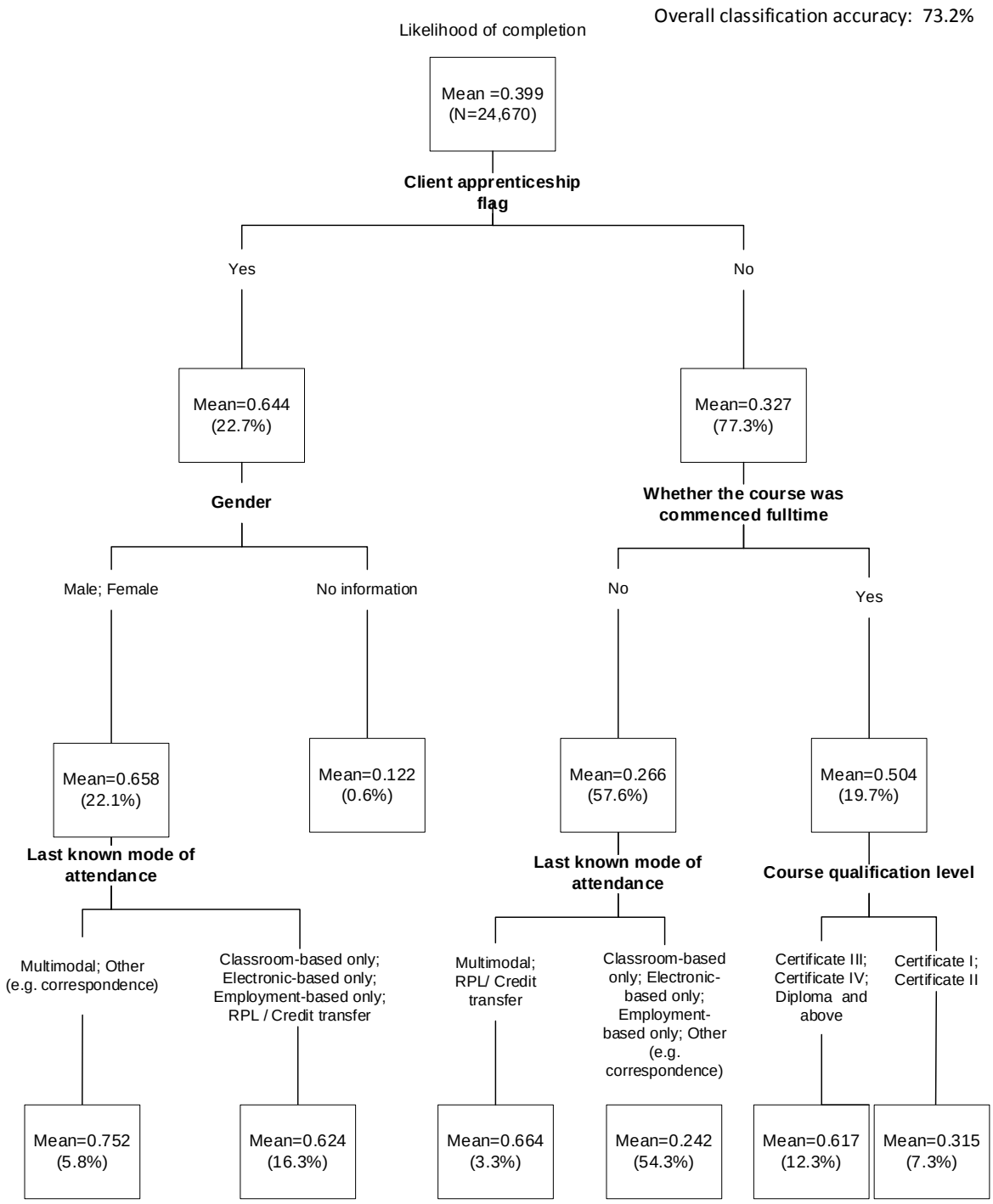
Note: Mean refers to the likelihood (i.e. probability) of completing a government-funded VET qualification. The percentage figure inside the parenthesis refers to the cluster size relative to the population frame in scope (i.e. N).

Figure C5 Decision tree diagram for Western Australia 2011 cohort



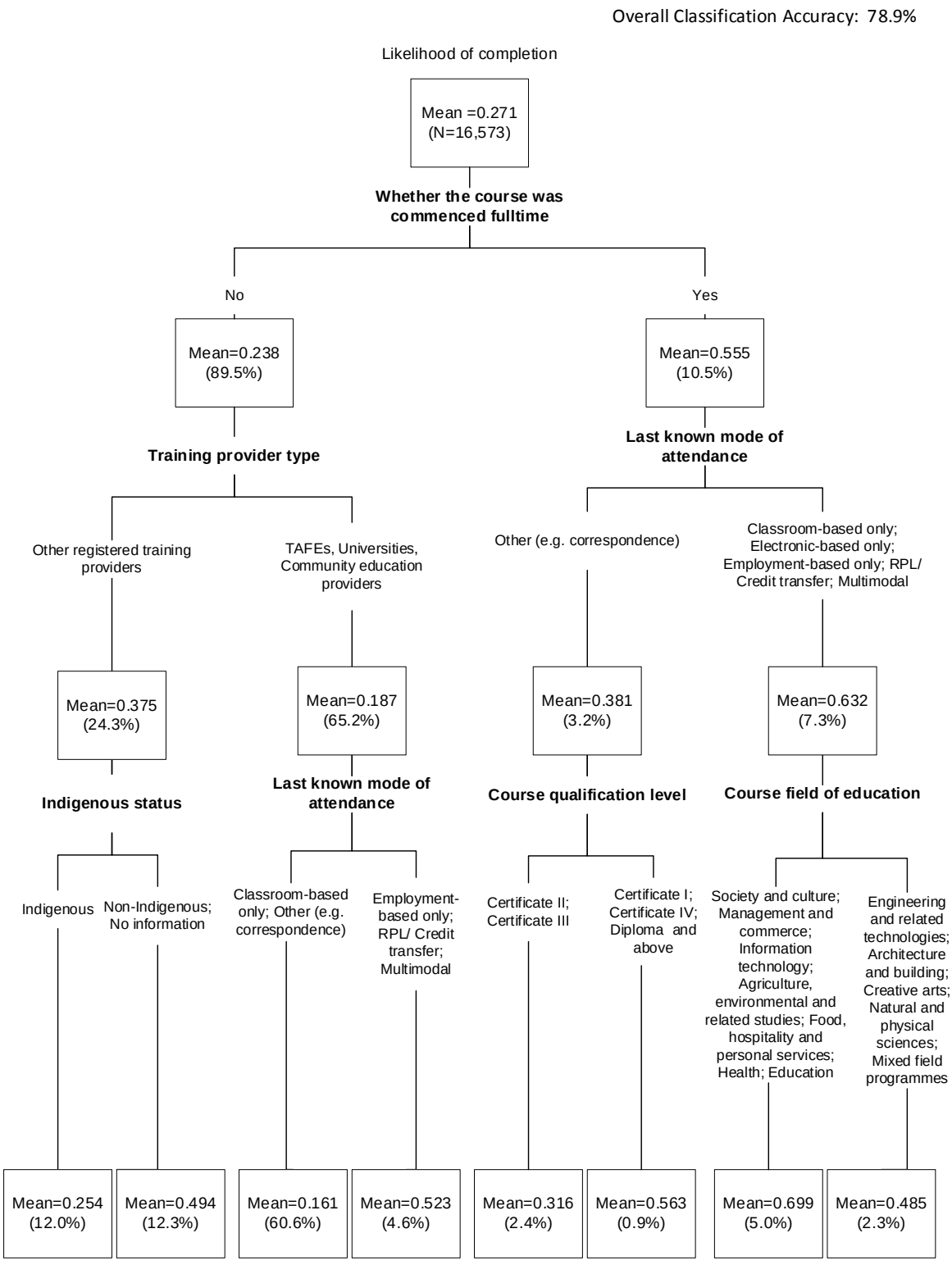
Note: Mean refers to the likelihood (i.e. probability) of completing a government-funded VET qualification. The percentage figure inside the parenthesis refers to the cluster size relative to the population frame in scope (i.e. N).

Figure C6 Decision tree diagram for Tasmania 2011 cohort



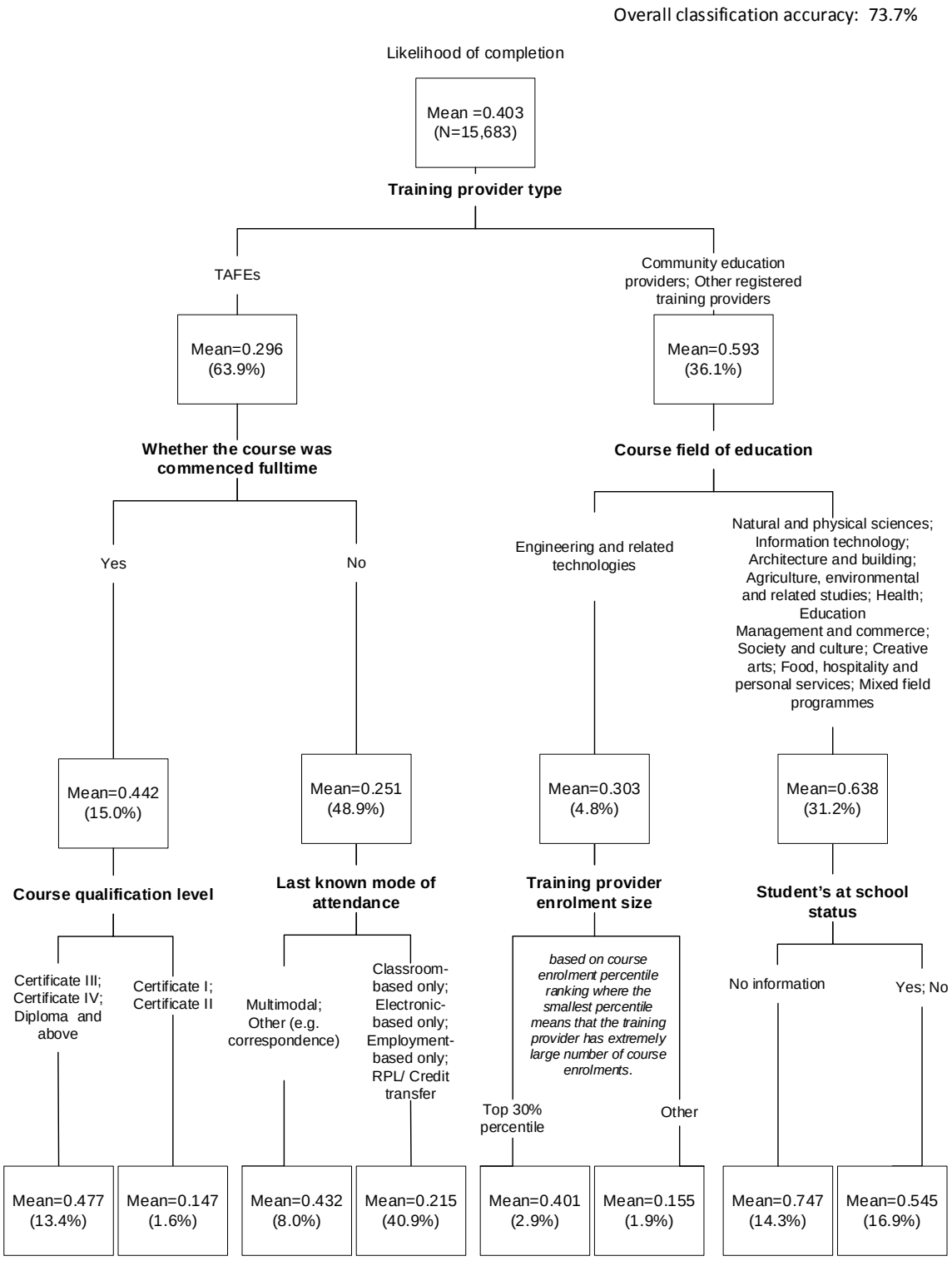
Note: Mean refers to the likelihood (i.e. probability) of completing a government-funded VET qualification. The percentage figure inside the parenthesis refers to the cluster size relative to the population frame in scope (i.e. N).

Figure C7 Decision tree diagram for Northern Territory 2011 cohort



Note: Mean refers to the likelihood (i.e. probability) of completing a government-funded VET qualification. The percentage figure inside the parenthesis refers to the cluster size relative to the population frame in scope (i.e. N).

Figure C8 Decision tree diagram for Australian Capital Territory 2011 cohort



Note: Mean refers to the likelihood (i.e. probability) of completing a government-funded VET qualification. The percentage figure inside the parenthesis refers to the cluster size relative to the population frame in scope (i.e. N).

National Centre for Vocational Education Research

Level 5, 60 Light Square, Adelaide, SA 5000
PO Box 8288 Station Arcade, Adelaide SA 5000, Australia

Phone +61 8 8230 8400 **Email** ncver@ncver.edu.au
Web <https://www.ncver.edu.au> <https://www.lsay.edu.au>
Follow us: <https://twitter.com/ncver> <https://www.linkedin.com/company/ncver>