

RESEARCH ARTICLE

## Profiling Students at Risk of Dropout at a University in South Africa

Ratoeba Piet Ntema\*

### *Abstract*

*Student dropout is a significant concern for university administrators, students and other stakeholders. Dropout is recognised as highly complex due to its multi-causality, which is expressed in the existing relationship in its explanatory variables associated with students, their socio-economic and academic conditions, and the characteristics of educational institutions. This article reports on a study that drew on university administrative data to build a profile of students at risk of dropout from 2008–2018. The study employed a data mining technique in which predictors were chosen based on their weight of evidence (WOE) and information value (IV). The selected predictors were then used to build a profile of students at risk of drop-out. The findings indicate that at-risk students fail more than four modules in a year with a participation average mark of 43% or less and have joined the university in the second academic year. It is suggested that universities put measures in place to control and prevent students who carry over four or more modules from adding modules to their registration until the failed modules are passed.*

### *Keywords*

*data mining, student dropout, weight of evidence, information value, risk profile*

### *Introduction*

Dropout rates in higher education are a significant concern in international and national contexts (Marquez-Vera et al., 2013; Orellana et al., 2020). The concept of dropout refers to the condition where students leave an academic programme either temporarily or permanently before the end of the academic year or before complying with the requisite requirements for graduation (Bonaldo & Pereira, 2016; Daniels, 2006; Letseka, 2007). According to the Organisation for Economic Co-operation and Development (OECD) the dropout rate increased in Australia, Austria, Belgium, Canada, Chile, Costa Rica, Colombia, Denmark, Estonia, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovenia, Spain, Switzerland, Turkiye, the United Kingdom and United States from 35% in 2005 to 64.5% in 2018 (Guzmán et al., 2021). The dropout rates also increased substantially in countries such as Luxembourg, Hungary, Sweden, the Czech Republic, and Slovakia (Guzmán et al., 2021). Another

\* Ratoeba Piet Ntema is a lecturer at North-West University, South Africa. Email: Piet.Ntema@nwu.ac.za. ORCID: 0000-0002-3379-3532.

example is the situation in Latin America, where dropout rates in higher education have historically been high, hovering around 54%, and are predicted to rise in coming years (Becerra et al., 2020). Due to its multiple causes, and the subsequent effects on various stakeholders, such as students, their families, higher education institutions (HEIs) and the broader society, dropout rates are also considered a major concern in South Africa (Mthalande et al., 2021).

According to Moeketsi and Maile (2008) in a report for the Human Sciences Research Council, the Department of Education predicted in 2005 that 36,000 (30%) of the 120,000 students who started higher education in 2000 would dropout during their first year. In their second and third years, another 24,000 (20%) left. During the three-year span, just 22% of the remaining 60,000 obtained their bachelor's degrees. According to the study, dropout rates at some universities may surpass 80%. Between 2000 and 2004, one out of every three university students and one out of every two Technikon students were predicted as likely to dropout. Nearly 20 years later, in 2020 HEI students' academic stressors were compounded (Crawford et al., 2020), partly as a result of the rapid and drastic transitions in higher education teaching and learning compelled by the outbreak of the COVID-19 pandemic. The majority of South African HE students, notably those from historically black universities and HEIs, were affected, resulting in a high dropout rate (Camilleri, 2021).

Previous research has attempted to identify the factors that explain current dropout rates and the reasons for high dropout rates (Camilleri, 2021; Mthalande et al., 2021; Moodley & Singh, 2015). Amongst others, researchers speculate that incorrect career choice, inadequate academic support, insufficient funding, relations with other students, stress factors such as accommodation issues, background of students (including families and finances), individual traits, pre-university (academic potential), challenges associated with the coronavirus (COVID-19) pandemic, and proficiency in the medium of instruction which some students struggle to cope with, as it affects their reading and processing skills, contribute to student dropout. This work has led to the development of tools and various perspectives that give decision-makers a comprehensive understanding of dropout prevention and mitigation (Kehm et al., 2019).

Notwithstanding previous international and national research, few studies have considered student administrative data in relation to dropout rates. Moreover, limited, if any, studies have reported on the use of statistical methods such as data mining to explore factors associated with dropout among South African universities. Instead, the majority of research into dropout has utilised primary response methods. Whereas primary response methods can provide relevant data, I hope to show here that other statistical methods, such as data mining, could offer unique insights into factors related to dropout within a South African context. Amongst others, a data mining approach could be used to build a profile of students at risk of dropout. Consequently, this article reports on a study that applied data mining techniques to administrative data to build a profile for students at risk of dropout from a university in South Africa.

The article begins with an overview of the literature related to the concept of dropout. Then, the methodology that guided the study is presented. This is followed by a presentation and discussion of the results. The article concludes with a summary of the main findings and their implications.

### *Review of the Literature*

Higher education is an enabler of life chances and research indicates that graduates are less likely to be unemployed compared to persons who did not obtain a post-school education (Scott et al., 2007). Higher education also has a direct bearing on women's employment opportunities, productivity growth, and entrepreneurship. It is a crucial element of socio-economic development (Latif et al., 2015; Pouris & Inglesi-Lotz, 2014). Thus, student dropout is not only a major concern for HEI administrators but can also result in various negative consequences for students, their families and the broader society (Cloete, 2014; Van der Merwe, 2020).

In terms of economic costs, Magnum et al. (2005) assert that student dropout has a negative effect on the financial management of higher education institutions. Amongst others, universities invest financial resources in student recruitment, teaching and learning, and accompanying student development and support initiatives (Paura et al., 2017; Ameri et al., 2016).

Dropout is costly to students as well. They lose earning potential and find themselves with immediate out-of-pocket expenses (Paura & Arhipova, 2014). According to Rincón et al. (2022), the student's dropout represents a sunk cost for the family because the costs incurred to pay for the studies were never recouped. It also represents the destruction or impossibility of creating long-term social capital that would have allowed the family to improve its socio-economic and educational conditions in the future (Ghignoni, 2017). The National Student Financial Aid Scheme (NSFAS) review noted that the 2010 data indicated that 48% of NSFAS-funded students had dropped out or not completed their studies (Breetzke & Hedding, 2016). This implies that students' dropout also has negative results on public funds.

Various researchers have attempted to identify risk factors related to student dropout (Aldowah et al., 2020; Hegde & Prageeth, 2018; St. John et al., 2000). *Inter alia*, the following have been identified as risk factors: behavioural problems, poor attendance, low socio-economic status, choice of institution, poor grades, and attendance with large numbers of poor students (Aldowah et al., 2020; Hegde & Prageeth, 2018; St. John et al., 2000). In addition to the aforementioned risk factors, Tinto's (1975) student integration model theories postulate that the interaction between students and the institution ultimately affects a student's decision to persist or not.

Although substantial studies have been conducted on student dropout, most have relied on primary response data methods. There are disadvantages associated with primary response data methods, such as the cost and time to develop resources involved in preparing the data, collecting a relevant data set, and managing the information;

feasibility and accessibility of enough participants and lastly, the risk of inaccurate feedback from participants (Wilcox et al., 2012).

Therefore, this study proposes using data mining techniques to look at the issues that may contribute towards student dropout. Data mining techniques can identify and predict future trends, track the behaviours and habits of participants and, lastly, assist with decision-making (Hsu & Yeh, 2019). In particular, the focus of the study was on profiling students at risk of dropout using administrative data obtained from a university in South Africa.

## *Research Methodology and Approach*

### **Data mining methods for student profiling**

This section describes the data mining methods used to profile students at risk of drop-out. In particular, the study used weight of evidence (WOE) and information value (IV) to profile students at risk of dropout. These strategies help to explore data and screen variables. The underlying theory of WOE was provided by Good (1950), and the expression describes whether the evidence in favour or against some hypothesis is more or less strong. Although frequently employed in scientific and social science research, WOE analysis is rarely used in education research (Weed, 2005). It calculates the percentage of events vs nonevents for a given attribute (Good, 1950). An event stands for something that has already happened, such as a student's dropout from university, and a nonevent represents the opposite, non-dropout.

### *Weight of evidence and information value*

Two data mining strategies for variable transformation and selection are the weight of evidence (WOE) and information value (IV). Because of the logarithm transformation used in WOE, they have a strong connection to logistic regression modelling, and IV is one of the most used feature selection methods when employing a logistic regression classifier (Zdravevski et al., 2011). The use of WOE involves a transformation of data that requires binning, which is a process that transforms a continuous or a categorical variable into set groups or bins. To initiate analysis, there is a need to assess the strength of each characteristic using the following criteria:

- The predictive power of each attribute is measured by the weight of evidence (WOE).
- The range and trend of WOE across attributes within a characteristic.
- The predictive power of characteristic is measured by the information value (IV).

The calculation process is carried out as follows. Let  $Y$  be a binary dependent variable and a set of predictive variables  $\chi_1, \dots, \chi_n$ . WOE can be used to measure the predictive strength of  $\chi_j$  and help to separate cases when  $Y = 1$  (dropout) from cases when  $Y = 0$  (non-dropout). The weight of evidence (WOE) method assists in converting a continuous independent variable into a set of groups or bins.

If  $\beta_1, \dots, \beta_k$  denote the bins for  $\chi_j$ , the WOE for  $\chi_j$  for bin  $i$  can be written as

$$WOE = \log \frac{P(\chi_j \in \beta_i | Y=1)}{P(\chi_j \in \beta_i | Y=0)} \tag{2.1}$$

To determine the IV for variable  $\chi_j$ , WOE is used as follows:

$$IV = \sum_{i=1}^k [P(\chi_j \in \beta_i | Y=1) - P(\chi_j \in \beta_i | Y=0)] \times WOE \tag{2.2}$$

Generally, if  $IV < 0.05$  the variable has very little predictive power and will not add any meaningful predictive power to a model. Table 1 summarises the criteria that can be used to interpret IV (Zdravevski et al., 2011).

**Table 1: Information value interpretations**

Information value (IV)	Variable’s Predictive Power
<0.02	Not useful for prediction
0.02 – 0.1	Weak predictive power
0.1 – 0.3	Medium predictive power
>0.3	Strong predictive power

When employing the WOE, the following eight empirical guidelines should be followed:

1. Each category should have at least 5% of the data.
2. Each category should be non-zero for both “dropout” and “non-dropout” observations.
3. The WOE for each category should be different.
4. Similar groups should be grouped together.
5. The WOE for non-missing values should be monotonic, going from negative to positive (or positive to negative) with no reversals.
6. Missing values should be binned separately.
7. The relevant weight indicates where the lost data categories/bins originate.
8. Experimenting with different categories will usually result in good student profiling.

Using WOE and IV has several advantages. First, nonlinear data transformation via WOE grouping greatly boosts a model’s flexibility in dealing with complex data patterns. Second, IV variable selection eliminates variables with low predictive power from the model, leaving only informative variables. Third, there are no restrictions on the input variable type (numerical or categorical), therefore, a variable’s scale (or unit) has no bearing on the modelling outcomes.

However, there are two distinct disadvantages to using WOE and IV. First, the binning method may result in information loss (variation). Second, no consideration is

given to correlations between the independent variables. For example, some independent variables may have a strong link, highlighting the significance of data exploration prior to implementing the approach.

### **Population and data sampling**

The population for the study consisted of all North-West University (NWU) contact (full-time) undergraduate students. The sample used includes student information spanning a 10-year period, from 2008 to 2018, made available for the study through request via ethical processes.

### **Data collection**

Two sets of data were used to identify factors that contribute to student dropout accurately. The first data set on student dropout rates was obtained from NWU's higher education management information system (HEMIS). The HEMIS tracks the student dropout rate through cohort studies using the students' unique student numbers. The second data set was the Programme Qualification Mix (PQM) at NWU. The PQM contains all the information about the institution's current qualifications.

To obtain the data, the study first went through the ethics clearance process of the university (ethics reference number: NWU-01271-19-S9). Student data were handled with care and no student was identified in the study. Names and university numbers were excluded, and new and unique ID numbers were assigned to data entries relevant to the selected period for the purpose of valid analysis. To protect the integrity and digital security of the data, the researcher created password-protected data files.

### **Research procedure**

This section discusses the pre-processing steps that were implemented in building the profile of the students at risk of dropout. Obtaining reliable and statistically valid data is crucial for the development of the profile of at-risk students. Therefore, the quantity and quality of data should comply with the requirements of statistical significance and randomness. Below, are the steps followed to ensure that the data were relevant for developing the profile (Siddiqi, 2012).

#### *Step 1: Definition of event (dropout)*

Dropout is defined as the interruption of studies by higher education students regularly enrolled for any length of time, regardless of university changes, before the conclusion of their study programmes (Bonaldo & Pereira, 2016).

#### *Step 2: Dealing with missing values*

The mode of the variable usually fills in the missing value of the data. The mode filling concept, which is based on the maximum probability filling approach and can improve the efficiency of data set integration, is aimed at the value with the highest number of

occurrences in the data. The missing values in the data were replaced by the means of the variables. If the missing value in the variable exceeds 95%, the variable is discarded.

### *Step 3: Checking correlation*

In the case of correlated variables, one variable from the correlated group of variables will be selected. The ideal variable will be the one that will theoretically represent all the information contained in the other variables of the group.

### *Step 4: Bucketing of the variables*

WOE was used to transform continuous independent variables into bins based on their similarities, whereby each bin contained more than 5% of observations. Furthermore, those bins did not have zero dropout nor non-dropout. After binning, WOE was calculated for every category as shown in equation 2.1. The calculated WOE was then used to calculate IV. The two concepts were then used to benchmark, screen, select and rank more suitable variables to predict the target variable by using their predictive powers. The criteria in Table 1 were used to select the variable with suitable predictive powers.

### *Step 5: Selection of variables*

Variables were pre-selected for the process to be efficient. The chosen variables were selected based on their predictive ability using WOE and IV. Weak variables were discarded in building a profile.

### *Step 6: Risk profile*

Finally, the process's main objective was to build a comprehensive risk profile for students at risk of dropout. The results of the process are presented in the next section.

## **Data description and analysis**

This section presents the description of the data used and the inclusion and exclusion of specific variables as part of data analysis.

### *Data description*

The total number of entries for this study was 495,771, with 28 columns as potential predictors of dropout. The data contained student information such as matric admission point scores (APS), personal demographics, university academic record, bursary information, residence status, duration of the qualification. For each of the 495,771 entries over the period 2008 to 2018, the study defined the binary dependent variable (dropout) as taking the value of 1 if the student dropped out and 0 otherwise.

### *Data analysis*

All analyses were conducted using Microsoft Excel and Python. The data set was then divided into two parts: training (0.8) and testing (0.2). Correlations between variables

were checked as mentioned in step 3 of the research procedure. Out of the 28 variables, 9 were correlated: module marks sum; module passed; module marks average; credits sum; passed count; qualification commencement year; exam average; matric average; and presentation method. For analysis, all correlated variables were removed. Table 2 shows the remaining 19 columns (variables) used for feature selection and profiling.

**Table 2: Variables used for feature selection and profiling**

Variables (Features)	Descriptions
First_Student_Year	Year of first registration to degree
Year_of_Birth	Year student was born
Gender_Eng	Gender of student
Entry_Level_Eng	Level at which the student joined the university
Undergraduate_Postgraduate_Eng	Undergrad or postgraduate identifier for the student
IP_Qualification_Type_2_Eng	Type of qualification student enrolled for
Qualification_Commencemnet_lag_Year	Number of years in a qualification
Qual_Minimum_Duration_in_Years	Minimum duration of the qualification
Graduated	Describes whether the student graduated or not
Enrolment_Count	Number of times student enrolled for the course
Metric_no_of_subjects	Number of subjects student had at grade 12
APS_Score; Matric_Avg	Admission Point Score (APS) and average marks in matric
Bursary	Indicator for bursary holder or not
Residence	Indicator for staying in university residence or not
No_of_modules	Number of modules student enrolled for in a particular year
Modules_failed	Number of modules student failed in a particular year
Modules_otherreasons	Other reason other than pass or fail
Terminated_Studies	Indicator for drop-out or not
Participations	Participation marks average

Weight of evidence and information value technique was used on the remaining 19 variables to select more suitable predictors according to their weights and information value.

### ***Results and Discussion***

This section presents the results of the process of profiling students at risk of dropout using weight of evidence and information value. Data analysis was implemented using Python scripts (Jupyter Notebook).

**Target variable distribution**

Of the 495,771 entries in the data, 478,477 were recorded as retained (non-dropout) and 17,294 as dropout (i.e. terminated studies) (see Figure 1).

**Distribution of Terminated Studies**

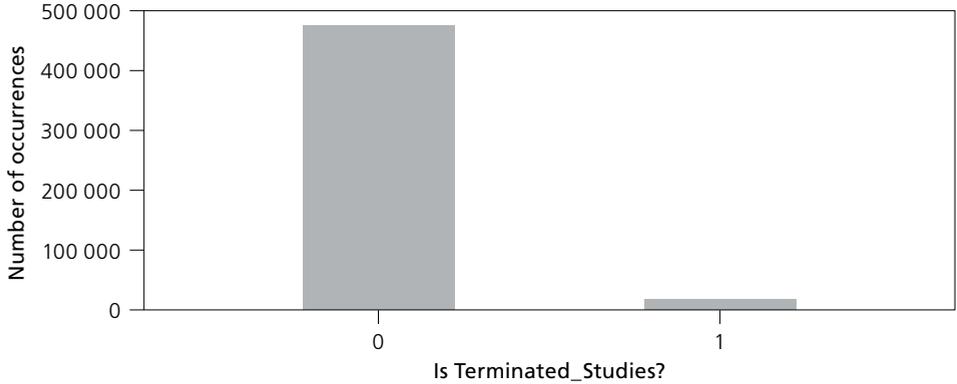


Figure 1: Distribution of the target variable

A further insight into student dropout in relation to the number of modules failed (Figure 2) highlighted that the percentage of student dropout increased sharply when students failed more than four modules.

**% Terminated\_studies & Acct Distribution: modules\_failed**

# Students — % Terminated\_studies

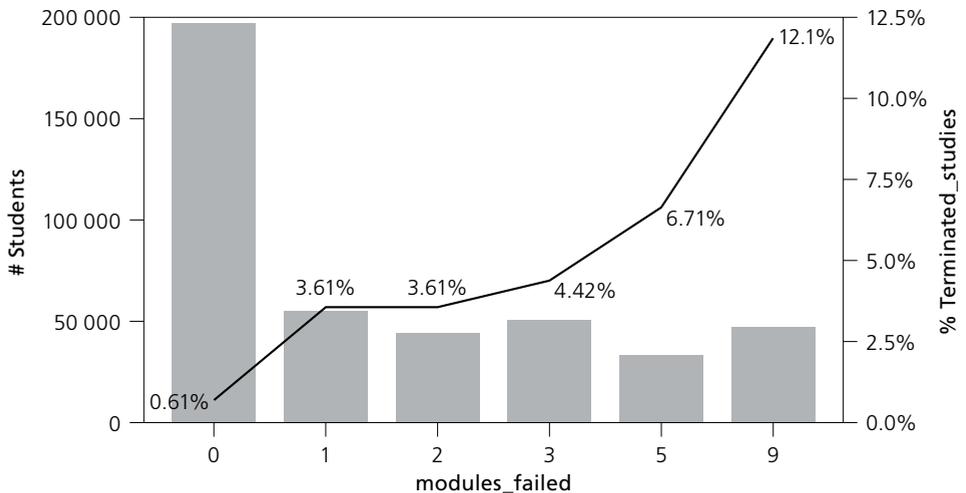
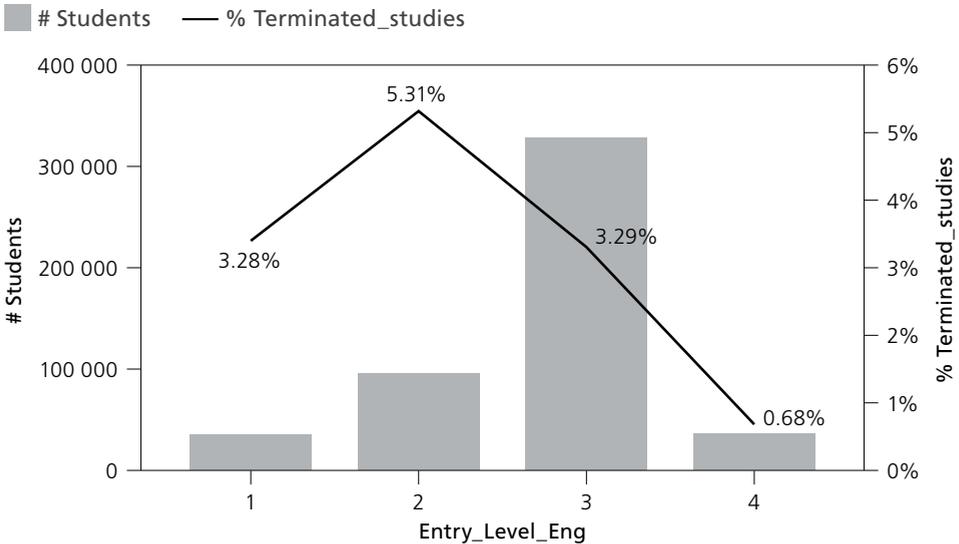


Figure 2: Distribution of students dropout in relation to number of modules failed

Figure 3 presents the distribution of dropout in relation to entry level at the university. Figure 3 shows a high dropout rate of students who entered the university at second-year entry level.

**% Terminated\_studies & Acct Distribution: Entry\_Level\_Eng**



**Figure 3: Distribution of student dropout in relation to entry level**

This section presents the results of the predictors that were used to build the profile of the students at risk of dropout. Tables 3, 4, 5 and 6 present the weight of evidence and information value of each predictor.

**Table 3: Weight of evidence and information value for entry level**

Cut Off	N	Events	%Events	Non-Events	%Non-Events	WOE	IV
1	35,061	1,151	0.07	33,910	0.07	-0.06	0.00
2	95,685	5,085	0.29	90,600	0.19	0.44	0.05
3	328,450	10,809	0.63	317,641	0.66	-0.06	0.00
4	36,575	249	0.01	36,326	.08	-1.66	0.10
	495,771	17,294	1	478,477	1	-1.346	0.15

The results in Table 3 show that a high proportion (29%) of event (dropouts) occur at entry level 2 as compared to (19%) of non-events (non-dropouts). At level 3, the (63%) proportion of events (dropout) is less compared to (66%) proportion of non-events (non-dropouts). However, the WOE for level 3 entry (-0.060) is less than level 2 entry, which implies there are more non-events (non-dropouts) at level 3 compared to level 2. Analysing the results in Table 3, the study can conclude that the WOE has more weight for entry level 2, and this suggests that students entering university in their second

academic year are likely to dropout as compared to other entry levels. According to the rule of thumb described in Table 1, the predictor's IV (0.15) indicates that it has a medium predictive value. Which implies that entry level 2 has medium predictive power to predict dropout of students.

**Table 4: Weight of evidence and information value for number of modules failed**

Cut Off	N	Events	%Events	Non-Events	%Non-Events	WOE	IV
(-0.01, 1.0]	293,925	3,723	0.23	290,202	0.61	-1.04	0.41
(1.0, 2.0]	42,769	1,541	0.09	41,228	0.09	0.03	0.00
(2.0, 3.0]	40,340	1,558	0.09	38,782	0.08	0.12	0.00
(3.0, 4.0]	26,091	1,319	0.08	24,772	0.05	0.39	0.01
(4.0, 7.0]	52,021	3,701	0.21	48,320	0.10	0.75	0.09
(7.0, 35.0]	40,625	5,452	0.32	35,173	0.07	1.46	0.35
	495,771	17294	1	478,477	1	1.71	0.86

The results in Table 4 show that a high proportion (32%) of events (dropouts) occur at the interval (7, 35] of modules failed followed by interval (4, 7] of modules failed with a proportion of (22%). From the results in Table 4, the study can conclude that the WOE has more weight for intervals (7, 35] and (4, 7] of modules failed than other intervals. This suggests that students failing more than four modules in an academic year are likely to dropout as compared to other students. According to the rule of thumb in Table 1, the predictor's IV (0.86) indicates a high predictive power. Which implies that number of modules failed has high predictive powers to predict students dropout.

**Table 5: Weight of evidence and information value for participation average mark variable**

Cut Off	N	Events	%Events	Non-Events	%Non-Events	WOE	IV
[0, 30]	49,679	6,930	0.40	42,749	0.09	1.50	0.47
(30, 43]	49,482	2,698	0.16	46,784	0.10	0.47	0.03
(43, 51]	57,542	1,927	0.11	55,615	0.12	-0.04	0.00
(51, 55]	44,306	1,093	0.06	43,213	0.09	-0.36	0.01
(55, 58]	49,903	1,176	0.07	48,727	0.10	-0.40	0.01
(58, 62]	60,725	1,194	0.07	59,531	0.12	-0.59	0.03
(62, 65]	44,691	676	0.04	44,015	0.09	-0.86	0.05
(65, 68]	41,984	580	0.03	41,404	0.09	-0.95	0.05
(68, 73]	50,926	570	0.03	50,356	0.11	-1.16	0.08
(73, 100]	46,533	450	0.03	46,083	0.10	-1.31	0.09
	495,771	17,294	1	478,477	1	-3.7	0.82

The results in Table 5 show that a high proportion (40%) of events (dropouts) occur in the interval [0, 30] of participation average marks which have (9%) of non-events

(non-dropouts). Another high proportion (16%) of events (dropout) occur in the interval (30, 43], which has (10%) of non-events (non-dropouts). Analysing the results in Table 5, the study can conclude that the WOE has more weight for intervals [0, 30] and (30, 43] of participation average marks than other intervals. This suggests that students obtaining a participation average mark (43%) or less in their modules for an academic year are more likely to dropout as compared to other students. According to the rule of thumb presented in Table 1, the predictor's IV (0.82) indicates that it has a high predictive power. Which implies that participation average marks have high predictive power to predict dropout of students.

### Profile of at-risk student

Table 7 presents a summary of the most suitable predictors of student dropout and their information value scores. According to the criteria in Table 1, entry level has medium predictive power, while the number of modules failed, and participation average marks have strong predictive power.

**Table 7: Summary of most suitable predictors of dropout**

Variables	IV	IV rank
Modules failed	0.86	1
Participation average	0.82	2
Entry level	0.15	3

According to the results presented in the previous section, the study can now profile students at risk of dropout as follows.

1. The student fails more than four modules per academic year.
2. The student obtains a participation average mark of 43 per cent or less.
3. The student has entered at second-year entry level.

### *Study Limitations and Further Research*

This results from this study should be read in light of certain limitations. First, the study only provides evidence that the variables described above may be relevant for at-risk student profiling for the administrative data utilised in this study but may not necessarily be exhaustive variables for profiling at-risk students in general. For example, there could be other relevant variables from qualitative data that are linked to, *inter alia*, student behaviour and attitude, university resources, and university leadership that were not considered in the analysis. Second, data were only collected from a single South African university. Hence, the external validity of the findings is limited.

Future research could focus on incorporating data from various university databases to develop a more holistic understanding of dropout rates among South African students. Furthermore, this study suggests that variables related to student behaviour and attitudes, university resources, university leadership, abilities, and skills of personnel (lecturers), teaching and learning environment, parental role, social aspects, health and psychological

issues, encouragement and motivation of students, study skills, time management, and other factors be included in profiling at-risk students for future research.

### *Conclusion and Recommendations*

This research aimed to use data mining techniques on university administrative data from a university in South Africa to create a profile of students at risk of dropout. Not all students will achieve their academic goals, and some will be labelled as at-risk. The risk profile may assist the university in identifying such students. After successfully identifying at-risk students, university officials and other university representatives may be able to establish appropriate intervention tactics and support programmes to help students at risk of dropout.

The study used WOE and IV to select suitable predictors with predictive power. To create a profile of at-risk students, the selected predictors were analysed. The study reached the following conclusions based on the examination of chosen predictors.

First, based on the criteria in Table 1, this study concluded that a student who has failed more than four modules in an academic year with a participation average mark of 43% or less has a high likelihood of dropping out without finishing their studies. Second, based on the findings that a student who enters the university at the second-year entry level is more likely to dropout, this study concludes that students who have previously dropped out (from another institution) will most likely dropout again.

As indicated in the data analysis section, the number of modules for which a student is registered has a strong correlation with the number of modules failed. The researcher recommends that universities put in place measures to control and prevent students who carry over four or more modules from adding more modules to their registration until failed modules are completed. This will assist students in managing the number of modules registered and focusing on failed modules. Furthermore, the control mechanisms could boost the chances of students receiving high participation marks, resulting in a high chance of passing the modules. A further recommendation is for universities to note that students who may have not been identified as at-risk in the current academic year, may be at-risk the following academic year. Therefore, a continuous monitoring system is needed.

For future research this study suggests inclusion of variables linked to student behaviour and attitude, university resources, university leadership, abilities, and skills of personnel (lecturers), teaching and learning environment, parental figure(s), social aspects, health and psychological issues, encouragement and motivation of students, study skills, time management, etc., to be included in profiling at-risk students.

### *Conflict of Interest*

The author declares no conflict of interest.

## References

- Aldowah, H., Al-Samarraie, H., Alzahrani, A. I., & Alalwan, N. (2020). Factors affecting student dropout in MOOCs: A cause and effect decision-making model. *Journal of Computing in Higher Education*, 32(2), 429-454. <https://doi.org/10.1007/s12528-019-09241-y>.
- Ameri, S., Fard, M. J., Chinnam, R. B., & Reddy, C. K. (2016, October 24–28). Survival analysis based framework for early prediction of student dropouts. In *Proceedings of the 25th ACM international on conference on information and knowledge management* (pp. 903-912). ACM.
- Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, 12(2), 155-187. <https://doi.org/10.1007/BF00976194>.
- Becerra, M., Alonso, J. D., Frias, M., Angel-Urdinola, D., & Vergara, S. (2020). *Latin America and the Caribbean: Tertiary education*. World Bank.
- Bonaldo, L., & Pereira, L. N. (2016). Dropout: Demographic profile of Brazilian university students. *Procedia – Social and Behavioral Sciences*, 228, 138-143. <https://doi.org/10.1016/j.sbspro.2016.07.020>.
- Breetzke, G. D., & Hedding, D. W. (2016). The changing racial profile of academic staff at South African Higher Education Institutions (HEIs), 2005–2013. *Africa Education Review*, 13(2), 147-164. <https://doi.org/10.1080/18146627.2016.1224114>.
- Camilleri, M. A. (2021). Evaluating service quality and performance of higher education institutions: A systematic review and a post-COVID-19 outlook. *International Journal of Quality and Service Sciences*, 13(2), 268-281. DOI: 10.1108/IJQSS-03-2020-0034.
- Cloete, N. (2014). The South African higher education system: Performance and policy. *Studies in Higher Education*, 39(8), 1355-1368. <http://dx.doi.org/10.1080/03075079.2014.949533>.
- Crawford, J., Butler-Henderson, K., Rudolph, J., Malkawi, B., Glowatz, M., Burton, R., Magni, P. A., & Lam, S. (2020). COVID-19: 20 countries' higher education intra-period digital pedagogy responses. *Journal of Applied Learning & Teaching*, 3(1), 1-20. <https://doi.org/10.37074/jalt.2020.3.1.7>.
- Daniel, S. S., Walsh, A. K., Goldston, D. B., Arnold, E. M., Reboussin, B. A., & Wood, F. B. (2006). Suicidality, school drop-out, and reading problems among adolescents. *Journal of Learning Disabilities*, 39(6), 507-514. <https://doi.org/10.1177/00222194060390060301>.
- Ghignoni, E. (2017). Family background and university dropouts during the crisis: the case of Italy. *Higher Education*, 73(1), 127-151. <https://doi.org/10.1007/s10734-016-0004-1>.
- Good, I. J. (1950). *Probability and the weighing of evidence*. Charles Griffin.
- Guzmán, A., Barragán, S., & Cala Vitery, F. (2021). Dropout in rural higher education: A systematic review. *Frontiers in Education*, 6. <https://doi.org/10.3389/educ.2021.727833>
- Kehm, B. M., Larsen, M. R., & Sommersel, H. B. (2019). Student dropout from universities in Europe: A review of empirical literature. *The Hungarian Educational Research Journal*, 9(2), 147-164. <http://dx.doi.org/10.1556/063.9.2019.1.18>.
- Latif, A., Choudhary, A. I., & Hammayun, A. A. (2015). Economic effects of student dropouts: A comparative study. *Journal of Global Economics*, 3(2). DOI: 10.4172/2375-4389.1000137.
- Letseka, M. (2007). Why students leave: The problem of high university drop-out rates. *HSRC Review*, 5(3), 8-9.
- Letseka, M., & Maile, S. (2008). *High university drop-out rates: A threat to South Africa's future*. Human Sciences Research Council.
- Hegde, V., & Prageeth, P. P. (2018). Higher education student dropout prediction and analysis through educational data mining. In *2018 2nd International Conference on Inventive Systems and Control (ICISC)* (pp. 694-699). IEEE.

- Hsu, C. W., & Yeh, C. C. (2019). Mining the student dropout in higher education. *Journal of Testing and Evaluation*, 48(6), 4563-4575. <http://dx.doi.org/10.1520/JTE20180021>.
- Mangum, W. M., Baugher, D., Winch, J. K., & Varanelli, A. (2005). Longitudinal study of student drop-out from a business school. *Journal of Education for Business*, 80(4), 218-221. <https://doi.org/10.3200/JOEB.80.4.218-221>.
- Marquez-Vera, C., Morales, C. R., & Soto, S. V. (2013). Predicting school failure and dropout by using data mining techniques. In *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje* (Vol. 8, pp. 7-14). IEEE. DOI: 10.1109/rita.2013.2244695.
- McWhirter, J.J., McWhirter, B.T., McWhirter, E.H., & McWhirter, R.J. (2007). *At risk youth*. Brooks/Cole
- Moodley, P., & Singh, R. J. (2015). Addressing student dropout rates at South African universities. *Alternation* (17), 91-115.
- Mthalande, P. P., Agbenyegah, A. T., & Dlamini, B. I. (2021). Reflection on student drop-out against the backdrop of COVID-19 in the South African Educational context amongst marginalised groups of students. *African Sociological Review/Revue Africaine de Sociologie*, 25(1), 194-217.
- Murray, M. (2014). Factors affecting graduation and student dropout rates at the University of KwaZulu-Natal. *South African Journal of Science*, 110(11-12), 01-06. <http://dx.doi.org/10.1590/sajs.2014/20140008>.
- Orellana, D., Segovia, N., & Rodríguez Cánovas, B. (2020). El abandono estudiantil en programas de educación superior virtual: revisión de literatura. *Revista de la educación superior*, 49(194), 47-64.
- Paura, L., & Arhipova, I. (2014). Cause analysis of students' drop-out rate in higher education study program. *Procedia – Social and Behavioral Sciences*, 109, 1282-1286.
- Paura, L., Arhipova, I., & Vitols, G. (2017). Evaluation of student dropout rate and reasons during the study. In *INTED2017 Proceedings* (pp. 2233-2238). IATED.
- Pouris, A., & Inglesi-Lotz, R. (2014). The contribution of higher education institutions to the South African economy. *South African Journal of Science*, 110(3-4), 01-07.
- Rincón, A. G., Moreno, S. B., & Cala-Vitery, F. (2022). Rural population and COVID-19: A model for assessing the economic effects of drop-out in higher education. *Higher Education Dropout After COVID-19: New Strategies to Optimise Success*. <https://doi.org/10.3389/feduc.2021.812114>
- Sampaio, G. R. (2012). *Three essays on the economics of education*. University of Illinois at Urbana-Champaign.
- Scott, I., Yeld, N., & Hendry, J. (2007). A case for improving teaching and learning in South African higher education. *Higher Education monitor* no. 6. Centre for Higher Education Development, University of Cape Town.
- Siddiqi, N. (2012). *Credit risk scorecards: developing and implementing intelligent credit scoring* (Vol. 3). John Wiley & Sons.
- Slavin, R. E. (1989). *Effective programs for students at risk*. Allyn and Bacon.
- St. John, E., Cabrera, A., Nora, A., & Asker, E. 2000. Economic influence on persistence reconsidered: How can finance research inform the reconceptualisation of persistence models. In J. Braxton (Ed.), *Reworking the student departure puzzle* (pp. 29-47). University Press.
- Stratton, L. S., O'Toole, D. M., & Wetzel, J. N. (2008). A multinomial logit model of college stopout and dropout behavior. *Economics of Education Review*, 27(3), 319-331. <https://doi.org/10.1016/j.econedurev.2007.04.003>.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 89-125. <https://doi.org/10.3102/00346543045001089>.

- Van der Merwe, R. L., Groenewald, M. E., Venter, C., Scrimnger-Christian, C., & Bolofo, M. (2020). Relating student perceptions of readiness to student success: A case study of a mathematics module. *Heliyon*, 6(11), e05204. <https://doi.org/10.1016/j.heliyon.2020.e05204>.
- Weed, D. L. (2005). Weight of evidence: A review of concept and methods. *Risk Analysis: An International Journal*, 25(6), 1545-1557. <https://doi.org/10.1111/j.1539-6924.2005.00699.x>.
- Wilcox, A. B., Gallagher, K. D., Boden-Albala, B., & Bakken, S. R. (2012). Research data collection methods: From paper to tablet computers. *Medical Care*, S68-S73. DOI: 10.1097/MLR.0b013e318259c1e7.
- Zdravevski, E., Lameski, P., & Kulakov, A. (2011, July). Weight of evidence as a tool for attribute transformation in the preprocessing stage of supervised learning algorithms. In *The 2011 international joint conference on neural networks* (pp. 181-188). IEEE.

***How to cite:***

Ntema, R.P. (2022). Profiling students at risk of dropout at a university in South Africa. *Journal of Student Affairs in Africa*, 10(2), 179-194. DOI: 10.24085/jsaa.v10i2.4077