

PREDICTING ACADEMIC SUCCESS USING LOGISTIC REGRESSION ON BIOGRAPHICAL DATA

Innocent Mamvura Wits Business Intelligence Siyaphumelela Conference 2017 27-29 June 2017

Acknowledgements

- Wits Business Intelligence Services
- Wits DVC Academic
- Wits Academic Advisors

Data is the Key to Success

"We are not going to run out of data anytime soon. Its maybe the only resource that grows exponentially. Maybe every 1.5 years we are seeing data double, and much of that data is social data, data about ourselves" Andreas Weigend

Abstract

The University of Witwatersrand adopted the Seidman(2006) Retention model for student success to identify students who are most likely to fail first year based on the Admission point score, High School Marks. The university has two faculty advisors in each faculty providing interventions to students. This paper used a Binary Logistic technique to predict academic success using qualitative variables collected through a structured questionnaire by the university. IBM SPSS was used to analyze the data and Computer skills, library usage skills, First Generation, Financial Aid, Residence and Distance was found to be significant. The findings contribute greatly to the existing work in the field of students success and the university will now focus its resources to the most needy students.

Problem Statement

- Wits is collecting large amounts of data pertaining to student background, school information, financial background, level of competence in computers skills, library usage skills etc.
- The existing Student Early Warning System is based on only quantitative variables.
- Identifying at risk students using the existing system is resulting large numbers of first year students with less resources to assist these students.
- Greater shortage of resources.
- No Improvement in first year success rates.

Findings

- Distance to campus, Financial Aid, Residence, First Generation, Computer Skills, Library Usage is a predictor of academic success.
- Working part time is not a significant.
- The predictive accuracy of the Binary Logistic model is 81.2%.
- These findings will help with allocation alignment of resources where they are needed the most.

Independent and Dependent Variables

- Dependent Variable is Success with 1 and 0
- Independent Variables are First Generation, Distance, Residence, Financial Aid, Computer Skills, Library usage, Work part time

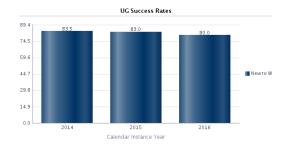
Definitions

 Binary Logistic Regression is a technique used when the outcome variable is a dichotomous variable (has two values). Logistic Regression uses Binomial Probability Theory in which there are only two outcome categories. The technique forms a function using the Maximum Likelihood Method, which maximizes the chances of grouping the observed data into the suitable category given the regression coefficients.

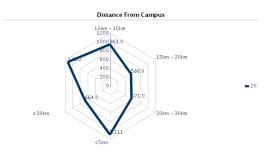
Assumptions of the Model

- Assumptions of Logistic Regression
- No Linear Relationship between the outcome and predictor variables is required;
- The outcome variable must have two categories;
- The predictor variables does not follow a normal distribution, or linear relationship;
- Maximum Likelihood coefficients are large sample estimates

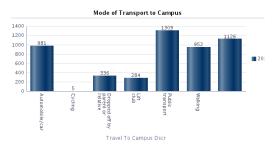
Wits UG Success Rates



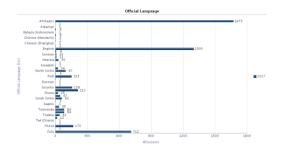
Data Visualization



Data Visualization



Data Visualization



Binary Logistic Model

		В	S.E.	Wald	df	Sig.	Exp(B)
	Work(1)	.038	.091	.170	1	.680	1.038
	Distance			33.641	5	.000	
	Distance(1)	.118	.122	.939	1	.332	1.12
	Distance(2)	.523	.125	17.470	1	.000	1.68
	Distance(3)	.594	.149	15.786	1	.000	1.81
	Distance(4)	.532	.142	13.939	1	.000	1.70
Step 1 ^a	Distance(5)	.067	.202	.109	1	.742	1.06
	Residence(1)	.629	.196	10.267	1	.001	1.87
	FinancialAid(1)	.622	.085	53.876	1	.000	1.86
	FirstGeneration(1)	.478	.084	32.245	1	.000	1.61
	LibraryUsage(1)	.206	.090	5.220	1	.022	1.22
	ComputerSkills(1)	.326	.093	12.350	1	.000	1.38
	Constant	.245	.108	5.189	1	.023	1.27

Model Implementation & Dashboards

												- 0/
	PS Score Report											
N 8	NG Support											
11.5	Internet 6/38/28/7 30 15:11 PM											
												61
		NAMES OF TAXABLE PARTY.			A CONTRACTOR OF THE						-	
		out order. The Presidence Lange	of some line of the				and and and and		- the second			and the second second
				The other	aport a intericted for th	te Assistant Depres						
				Dr.N. Act, shade	et sumber is about	statuted out own	siesents.					
												61
				a spaits years								
				Cargad St. J. S. MC21								
				aread to / a in the								
						IEAS PHILEER.) TH	EN 8 ESSE 3 END is equal to	NAME:				
				al 23 (a is MATRO								
					um - JEhrn, >354m							
				nic is easily to 7 is in								
				es la equé te / la le	N C							
		840	d Loud Sch Library i	sealto/ain#								
			d Deed Sch Library I d Dr. Residence is eq	negal to / is in M at to / it is Not to	Residence .							
			d Deed Sch Library I d Dr. Residence is eq	sealto/ain#	Residence .							
			d Deed Sch Library I d Dr. Residence is eq	negal to / is in M at to / it is Not to	Residence .							
1.7494			d Deed Sch Library I d Dr. Residence is eq	negal to / is in M at to / it is Not to	Residence .							47
	847.53	Ξ	d Deed Sch Library I d Dr. Residence is eq	n agust to / in in M pair to / in in Mait In gust to / in in Mait an	Residence .	-1				Provide Server	6	47
-		=	a baad Schuberge is oo a bi fooderson is oo a fee hat badan oo a fee hat badan oo	e equal to / is in M per to / is in Med Ta que to / is in Med ca conver - Tabul	Radidanca Francisk akt							
hagean (odo Kating Value Student Aue	and and and and and and and and and and	Deed Schubersy i 1 Headence is ea 1 Headence i	n egual to / in 11 M pai to / in 11 Met 18 gai to / in 11 Met 28 gaint to / in 14 Met 29 gaint to / in 14 Met 20 gaint to / in 14 Met 20	Radidanca Francisk akt					IT THE GIRD		Digital Seco
hagean (Colo: Rating Value: Student Num 20 - 204000	and and and and and and and and and and	Deed Schubersy i 1 Heidence is ea 1 Heidence i	e equal to / is in M per to / is in Med Ta que to / is in Med ca conver - Tabul	Radidanca Francisk akt		Ret on Inendel 40	Not In Residence	66	12 Page 612 6		tryist Secr
hagean (Colo: Rating Value Student Aum 20 (J74590 40 (HESS))	and and and and be School Name Thata gool High School Nagh angen and School Nagh angen and School	Louid Sch Library Lin Heiderson is eq Pro Not Makes on a Provide rad to Provide rad to Deat Promi Campo >330m	n nguai to / A in M pai to / A in Mart to cart to / a in Mart to care - Tatal a Daor Atan Unio At A	Radidanca Francisk akt		Not on Theredal and Not on Theredal and	I Not In Residence I Not In Residence	8 13	12 Plate 612 6		Daylor Seco
hagean (200 Kaling Weat Student Rue 27 (274590 40 (40103) 40 (10014)	and and and and and and and and and and	Dead Schuberge II 1. Novidence II no 1. Novidence II no 1. Novidence II no 1. Novidence II 2. State 1. State 1. State 1. State 1. State 1. State	n regulat to / A in 16 part to / A in Next To cont to / A in Next To cont to / A in Next To contex - Total a Date: Alem Lines Ad A A	Radidanca Francisk akt		Not on Trendel at Not on Trendel at Not on Trendel at	i Not In Residence I Not In Residence I Not In Residence	8 17 60	12 Mart 612 6	Prod. CC	Degled Seco
hagean (240 Kating Velue Student Num 25 (279/590 40 (269/33) 40 (269/34) 40 (269/34)	end and ter School Terres Thata goot might School School Companyones Rivers Calevisian sign School	4 Dead Set Library 1 4 Th Nexternal Is an President of S Deal Priori Carego > Xhen > Xhen > Xhen > Xhen > Xhen > Xhen > Xhen > Xhen	e equal to / is in M per to / is to Yest Ta cont to / is in Next Ta cont to / is in Next to Convert - Table in Decr Plane Univ. At in A	Radidanca Francisk akt		Ret on Trendel at Not on Frendel at Not on Frendel at Not on Trendel at Not on Trendel at	i Net In Residence Mat In Residence Mat In Residence Mat In Residence	81 10 10 14	12 Men 612 5		I biglish Second
hagean (200 Rating Veloc Student Rum 25 (274599) 40 (40503) 40 (40104) 40 (40104) 41 (201292)	end and end the School Nemit Thete pointings School Sophangere Cardioned School Andrer Gamptiones School Sand Caregoliensius School	4 Dead Set Udary) 8 In Residence is no 9 In Residence is no 9 In Residence is no 9 In Residence is 10 In Residence is 1	expand to / A 11 M per to / A 10 Meet To court to / A 10 Meet To court to / A 10 Meet To courts - Total a base / New Lives AD A A A A A A A A A A A A A A A A A	Feedbesch Franceist als ordensz über Sch D S S S S S S	Miguitaria Unal Sch N N N N	Net or Trender at Net or Trender at Net or Trender at Net or Trender at Net or Trender at	i Net In Residence Met In Residence Met In Residence Met In Residence Met In Residence	8 17 66 72	17 Mer 612 6	Prod. CC	I tryine Seco
Program (200 Fating Weild Student Rum 30 (2000) 40 (2000) 40 (2000) 40 (2000) 41 (2000) 41 (2000) 41 (2000)	end and the second seco	4 Dead Set Udaray 8 In Henderson Is ed 14 Not Roban Is ed 14 Not Roban Is ed 20 Deat Prom Campus > Xhem > Xhem 20 Am 20	e equal to / A in M arito / A in Mat La cont to / a in Mat La cont to / a in Mat or cont a ford a ford A A A A A A A A A A A A A A A A A A A	Radidanca Francisk akt		Not on Trence at Not on France an Not on France an Not on Trance at Not on Trance at Not on Trance at	i Net In Residence Mat In Residence Mat In Residence Mat In Residence Mat In Residence Mat In Residence	8 11 64 12 72	27467 622 6	Prod. CC	I tryine Seco
	200 Saley Veta States New 200 Saley Veta States New 40 (1003) 40 (1004) 40 (1003) 41 (2003) 41 (2003) 4	en e	4 Dead Set Udaray 9 Honderson is no 14 An Antibase is no 14 Antibase Campo + 20am + 20am - 20am 20am - 20am 20am - 20am 20am - 20am - 20am	e equal to / A in M per to / A in Met To cont to / A in Met To cont to / a in Met To cont of a in Met To cont of the in Met To A A A A A A A A A A A A A A A A A	Feedbesch Franceist als ordensz über Sch D S S S S S S	Miguitaria Unal Sch N N N N	Not on Trence at Not on Trence at	 Net In Residence Nat In Residence 	80 17 66 77 77 77	12 Men 612 6 73 80 64 62 73 61	Prod. CC	i biylidi Sece
Program (200 Fating Weild Student Rum 30 (2000) 40 (2000) 40 (2000) 40 (2000) 41 (2000) 41 (2000) 41 (2000)	end and the second seco	4 Dead Set Udaray 8 In Henderson Is ed 14 Not Roban Is ed 14 Not Roban Is ed 20 Deat Prom Campus > Xhem > Xhem 20 Am 20	e equal to / A in M arito / A in Mat La cont to / a in Mat La cont to / a in Mat or cont a ford a ford A A A A A A A A A A A A A A A A A A A	Feedbesch Franceist als ordensz über Sch D S S S S S S	Miguitaria Unal Sch N N N N	Not on Trence at Not on France an Not on France an Not on Trance at Not on Trance at Not on Trance at	Not in Residence Mat In Residence	8 11 64 12 72	27467 622 6	Prod. CC	I tryine Seco

Model Implementation & Dashboards

	/2017 10:28-36	int Report												
Het Table														8/3
		1 M/F - 10					2817 0				e conse li que il			
parter Association	Pogan	Student Number	First Name	Last Name	Campus Driel Address	Admission Point Score	Familitie Attentioner	Dist From Campus Decr	Unad Sch Computers	Used Sch	In Assistant	Fit Ad Status	Unit Ca	Unit Risk Flank
Any-Jul	58000	>062509	Bonge	Gougeta	1962509@etudents.wbs.ac.m		*	20km - 30km	N		Not 3n Residence	Not on financial aid	COM51115 PMTY0036	
		1375665	Ladiy	Peacled	1375885@vbudents.who.ac.an			23km - 38km	н	•	Not In Residence	Not on financial and	P\$67+0034 P\$67+0036 P\$67+0036	37.00
		1379637	Thebeng	Pargale	1379637@students wits ac as	•		>itten	N	*	Not In Residence	Not on Strancial aid	P907+0036	44.00
		1390993	Achanane	Nosi	136393@dudetts wits at an			20km - 30km	N	*	Not In Residence	Not on financial aid		
		1293179	Immeculate	Pilleri	1303179@vbudents.wb.ac.pr			>30em	N		Not 3n Residence	Not on financial and	P9(710041	33.0
		2405525	Vincent	Lakhele	145529-Avent with a car			-3844	м		Residence	Not on linancial and	P907H003H P907H0036	
		3434894	Lanegang	Galeng	104940 data altaca	•		20km - 20km	ы	•	Not In Residence	Not on Shancial aid	COMSISTS PROVIDER PROVIDER	38.0
		1435662	Mehlabane	Pendil	1406662@students.wks.ac.m			>30en	N	8	Not In Residence	Not on financial eld	P\$4710034	
		3427273	Tahopang	Pletomi	10777)@dudets.ets.et.or		N.	20km - 30km	м	*	Not 3n Residence	Not on Shanelal and		
		2477629	Kgatiso	Plahwaraci	1479250 dudents with ac an			20km - 30km	N	*	Residence	Not on financial aid		
		1508148	Thabisle	Similare	188940 more with as an		N.	>liku	N		Not In Residence	Not on Shancial aid	000L1005 O4EN0312	
		1000003	Incontin	Photo	1977 Mondate and a re-			Zin - Xin			Bell In	Ref on Strendid	P96710041	

Future Work

- Include more variables from the BQ data
- Use other machine learning techniques
- Explore deep learning algorithms
- Put the power of data into the hands of students
- Collaborate with other Universities on BQ

Conclusion & Recommendations

- Students travelling long distances to campus, not on financial aid, first generation, not in residence, no computer skills, no library usage skills are most likely to fail their first year.
- Working Part time is not significant.
- The university could provide accommodation to students staying >20km from campus and also provide access to computers and workshops to the students who had not been exposed to computers at high school. The university could also perhaps provide studying skills workshops and general use of library resources to these students. Academic mentors could also be arranged for first generation students as they require motivation and guidance.

