Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 2, 2021, page 69-78

Universiti Teknologi Malaysia (UTM) Mathematics Graduates' Employability and Underemployment Investigation During Pandemic COVID-19

[1]Hanis Syafiqah Abd Razak and [2]Haliza Abd Rahman

[1,2]Department of Mathematical Sciences
Faculty of Science, Universiti Teknologi Malaysia,
81310 Johor Bahru, Johor, Malaysia.

e-mail: [1]hanis-1998@graduate.utm.my, [2]halizarahman@utm.my

Abstract This research investigates the Mathematics graduate from Universiti Teknologi Malaysia's employability by using logistic regression. COVID-19 has caused a surge in the unemployment rate. Furthermore, some employees have to endure more workloads or a decrease in income. This research's output may help Mathematics undergraduate to understand better and prepare when entering the workforce. This investigation uses the 2016 – 2021 Mathematics graduates' data. Factors significant to employability are demographic factors (gender, ethnicity, age and type of region reside) and educational background (CGPA and course program). For underemployment, income, type of organization worked for, and working experiences are included as factors.

Keywords logistic regression; COVID-19; graduates' employability and underemployment,

1 Introduction

The Coronavirus 2019, better known as COVID-19, was first reported at the end of 2019 in Wuhan, China.COVID-19 had spread so fast that in just four months, it has infected all over the world. On 11th March 2020, the World Health Organization (WHO) has announced COVID-19 as a pandemic[1].

With social distancing as the new norm, the world economic landscape had to flip to sustain itself. So due to this, numerous companies had to lay off a large number of their workforce. This devastating action had caused millions of people to lose their jobs and hence their source of income. The most affected groups are the low-income workers. The Employment Information and Analysis Service (EIAS) had reported that a staggering 107,024 workers in Malaysia lost their job in 2020. Incidentally, an increase of 167.0% people loss employment compared to the previous year [3].

2 Literature Review

Logistic regression is one of the generalized linear models. In some ways, it has many similarities to simple linear regression. Nonetheless, simple linear regression is limited to only modelling with continuous variable problems. In comparison, logistic regression is used when dealing with the dichotomous dependent variable. Furthermore, logistic regression is also part of machine learning utilised to model binary (0,1) situation-based problems [3]. Not only is capable for one-on-one problem but also can be used for more than one covariance factors..

2.1 Logistic Regression in Employability Research

When researching for employability, logistic regression is commonly used to model the problem. Notably, it is often dealing with categorical variables. For instance, the dependent variable would either employed, ‘1’ or not employed, ‘0’.

A 2019/2020 final year student had modelled employment suitability among UTM Faculty of Science graduates using logistic regression. The predictors are for the influence factor, gender, ethnicity, course, CGPA, MUET, and working while studying. The odd ratio further explained that female graduates of the Faculty of Science are 2.649 times less likely to have suitable jobs than their male counterparts. In addition, t other ethnicities are 8.054 less likely to have relevant employment than their Malay peers [4].

[5] had investigated the link between female undergraduates in a university-sponsored employment skills development program in the United Arab Emirates and employment post-graduation. They used logistic regression with the primary variable was the participation in the World of Work program during undergraduate study. Participants of the World of Work program are found to be 6.7 times more likely to secure employment in their post-graduate than their non-participant counterpart

2.2 Logistic Regression in COVID-19 Research

Recently, [6] have examined who are more likely at risk to lose income because of COVID-19 in China. The studied relation was education level, family economic status, Communist Party membership, employment sector and *hukou* status against the risk of income losses. The analysis contained 4,715 respondents, with a majority of them from Hubei. They inquired about any changes in the income of the respondents. This study found that people without high school education are twice likely to lose their income than tertiary education qualification. The same goes for individuals who came from low-income families, they are also twice likely to lose income than individuals from a wealthier family. The self-employed are more vulnerable to lose income with a probability of 0.29. As for the location factor, Hubei province residents are four times likely to lose income than the other parts of China

In Chile, [7] used multinomial logistic regression modelling to estimate the perceptions of individuals and companies when facing the COVID-19 pandemic in the Ñuble region. The study took 313 citizens and 51 companies to explore the perception of the effect of COVID-19 on the country’s economy. The variables used were sex, age and demographic. They found that women and the elderly are more vulnerable to job security.

3 Methodology

Data was collected using Google Form sent through e-mail to the the Mathematics alumni. The questions detailed in inquiring about their demographic information, educational and professional whereabouts.

3.1 Statistics Testing

3.1.1 T-Test

The t-test is to determine the relationship between CGPA and the employability of graduates. For the hypothesis assumptions:

H_0 : There is no correlation between CGPA and employability

H_1 : There is a correlation between CGPA and employability

Level of significance, α = 0.is used to test this hypothesis. The t-test value calculation uses the formula:

$$t_{test} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{1}$$

Where n is the number of samples and r is the correlation coefficient with the degree of freedom, $df = n - 1$. H_0 will be rejected if the p-value < 0.05.

3.1.2 Chi-Square Test

The Chi-Square test is to examine the significant relation correlation between two categorical variables.

H_0 : There is no correlation between employability and variable a

H_1 : There is a correlation between employability and variable a

Where a stand for gender, region and course program. Using level of significance, α = 0.05.

$$\chi^2 = \sum_{i}^{r}\sum_{j}^{c}\frac{\left(x_{ij} - Eij\right)^2}{E_{ij}} \tag{2}$$

where $E_{ij} = \frac{(n_i n_j)}{n}$

x_{ij} is number of observations

E_{ij} is expected frequency

r is number of rows

c is number of columns

n is number of size sample

Reject H_0 if calculated $\chi^2 < \chi^2_{0.05,df}$ where there is a correlation between the variables stated and employability.

3.2 Logistic Regression

The logistic model is a binary probability model. It is the most suitable method to use in dealing with characterized variable or when it binary-based. The variables in logistic regression are called predictors. The binary variable is modelled based on the response variable or the dependent variable. The logistic model assumptions can turn into transforming Prob{Y = 1} to make a model that is linear in βx:

$$\text{logit}\,\{Y = 1|X\} = \text{logit}\,(P) = \log\left[\frac{P}{1-P}\right] = \beta x \tag{1}$$

where βx can be written as $\beta x = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i$. For β_0 is the intercept and β_i, where $i > 0$ is the regression coefficient for x_i .

3.2.1 Odd Ratio Test

The odd ratio is used for events of employed, furthering studies, no income change and more workload because of COVID-19.

$$\text{Odd} = \frac{\text{Probability of event}}{\text{Probability of non-event}} = \frac{p}{1-p} \tag{4}$$

3.2.2 Confidence interval in Logistic Regression

Using 95% as the confidence interval, it is to estimate the true parameter, β within the range of α = 0.05 of the estimated parameter, $\hat{\beta}$.

$$\beta \pm z\alpha/_2(\text{S.E}\ \hat{\beta}) \tag{5}$$

3.2.3 Univariate Analysis

Each explanatory variable will be model in logistic regression singularly to the responding variable, *Y*. This analysis is to study the relation of each predictor to *Y* without interruption of other variables. The result from this analysis can help in formulating the model.

$$\hat{y} = \hat{\beta}x + c \tag{6}$$

3.2.4 Multivariate Analysis

Unlike univariate analysis, the multivariate analysis combined all predictors to study relation with the responding variable, *Y*. It is to study the correlation between all predictors with *Y*.

$$\hat{y} = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \cdots + \hat{\beta}_n x_n + c \tag{7}$$

3.3 Goodness of Fit Test for Logistic Regression

3.3.1 Likelihood Ratio (LR) Test

This test statistic is the ratio of the likelihood at the estimated parameter, $\hat{\beta}$ values to the likelihood of the data at the maximum Using $-2\times$ the log of this likelihood ratio to obtain a desirable statistical properties [9]. The likelihood ratio test statistic for the logistic regression model is the log-likelihood.

$$l(\beta; y, X) = \sum_{i=1}^{N}\left[-\ln\left(1 + exp\left(x_i\hat{\beta}\right)\right) + y_i x_i \hat{\beta}\right]x_i \tag{8}$$

3.3.2 Hosmer and Lemmeshow Test (HL)

The Hosmer and Lemmeshow (HL) statistics is an alternative to use of deviance

$$l(\beta; y, X) = \sum_{i=1}^{N}\frac{(O_i - E_i)^2}{E_j\left(1 - \frac{E_i}{n_i}\right)} \sim \chi^2 \tag{9}$$

Where:
n_i = number of observations
O_i = number of observed data
E_i = number of expected data

4 Results and Discussion

4.1 Descriptive Analysis

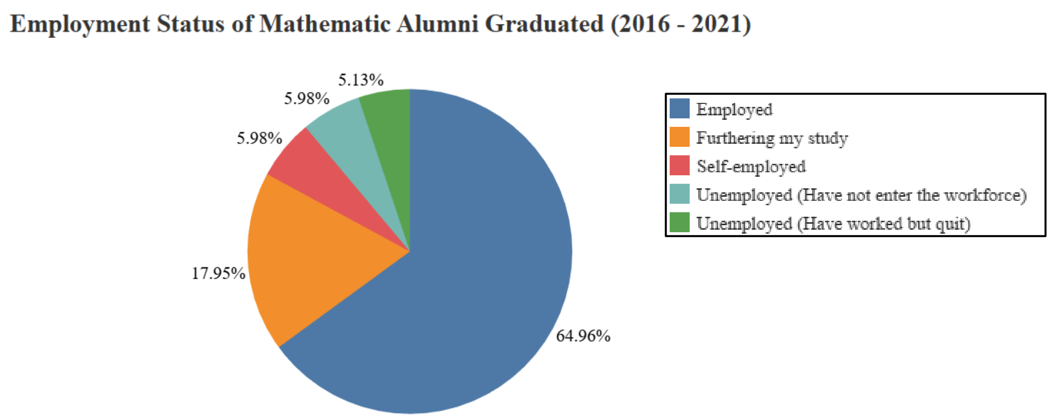

Figure 1: Pie Chart representation of the respondent's Employment Status

Based on Figure 1, the pie chart illustrates the total percentage of each employment status of the Mathematics alumni. A respondent who chose in training will be considered as employed. So, that would make a total of 76 mathematics graduates from 2016 till 2021 are in employment. In other means, 64.96% of the graduates are employed, 21 graduates (17.95%) are furthering their studies, seven alumni (5.98%) are self-employed and unemployed, and lastly, only six graduates(5.13%) are those who have worked but quit.

4.2 Statistics Testing

4.2.1 T-Test

Table 1: T-Test results

Variable	T-test Value	*p*-value
CGPA	1.094	0.256

The mean of CGPA was Mathematics graduate (2016 – 2021) is 3.40. From Table 1, the t-test value = 1.094 > *p*-value = 0.256. Because of that, H_0 is rejected. Hence, there is no correlation between CGPA and Mathematics Graduate Employability.

4.2.2 Chi-Square Test

Table 2: Chi-Square test results

Variables	Chi-square Value	df	*p*-value
Ethnicity	4.690	4	0.446
Gender	0.826	1	0.366
Region	18.090	1	0.055
Course Program	1.075	1	0.301

From Table 2, none of the variables tested has Chi-Square values less than the *p*-value. So, H_0 is rejected. Hence it can be concluded that no correlation between these variables to employability.

4.2 Predictive Analysis

4.2.1 Employment Prediction Analysis

Table 3: Univariate Analysis in Employability

Variables	Wald	df	*p*-value	Exp(*B*)
Demographic Background				
Malay	0.124	1	0.725	0.837
Chinese	1.036	1	0.309	2.294
Indian	2.431	1	0.119	0.250
Bumiputera Sabah	0.495	1	0.482	2.222
Bumiputera Sarawak	0.000	1	1.000	8.83×10^{9}
Gender	0.817	1	0.366	1.600
Age	**3.695**	**1**	**0.055***	**1.378**
Region	**16.774**	**1**	**0.000****	**5.641**
Education Background				
CGPA	1.289	1	0.256	0.529
Course Program	1.070	1	0.301	1.497

Note: $^{**}p < 0.05$, $^{*}p<0.10$

From Table 3, the univariate analysis shows that for the level of significance, $\alpha = 0.05$, the predictor region, x_4 and internship, x_7 has a significant *p*-value. For age, x_3 variable, it has very near to the significant value with only 0.005 difference.

In region, x_4 predictor, the Urban area is coded with '1' meanwhile Rural with '0'. The exp(*B*) value shows that graduates who live in the Urban zone are 5.641 at odds to their Rural counterparts in getting employed. Taking the log of odds, graduates who live in Urban are 1.73 times more likely to be employed than their Rural peers.

Table 4: Model Summary

Goodness Fit Tests	Enter Method	Forward	Backwards
-2 Log Likelihood	120.446	129.443	122.541
Cox & Snell R square	0.234	0.172	0.220
Nagelkerke R Square	0.322	0.237	0.303
Hosmer and Lemmeshow Test			
Chi-square	14.770	0.318	18.334
p-value	0.064	0.853	0.019

From Table 4, the -2 Log-Likelihood (-2LL) for all three methods have significant values, having greater than 100 each. Therefore, the three methods do not show excellent fitness. For each method, based on Nagelkerke's R-Square value, the Enter Method poses 32.2%, Forward 23.7% and Backward 30.3% of the variation in the outcome variable for the probability to be employed. Since all methods have below 50% Nagelkerke's R-Square value, all three approaches do not fit this dataset very well. All methods also have a *p*-value > 0.05 from the Hosmer and Lemmeshaw Test. The greater the *p*-value, the better the method matches to fit the model. However, for the forward method, it has a *p*-value = 0.853 > Chi-square = 0.318. This indicates that the forward method is poorly fit for the data [9]. Since the Enter method has a greater *p*-value = 0.064 than the Backwards method *p*-value = 0.019, the enter method will be used to model the data.

Table 5: Results of the Logistic Regression from SPSS

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(*B*)	95% C.I.for EXP(*B*)	
								Lower	Upper
Step 1[a]	Malay	-19.129	40192.337	.000	1	1.000	0.000	0.000	.
	Chinese	-17.381	40192.337	.000	1	1.000	0.000	0.000	.
	Indian	-20.817	40192.337	.000	1	1.000	0.000	0.000	.
	Sabah	-17.453	40192.337	.000	1	1.000	0.000	0.000	.
	Gender	0.884	0.692	1.631	1	0.202	2.421	0.623	9.404
	Age	0.070	0.191	.135	1	0.713	1.073	0.737	1.561
	Region	1.942	0.490	15.724	1	0.000	6.974	2.670	18.213
	CGPA	-1.585	0.764	4.303	1	0.038	0.205	0.046	0.916
	Course	0.005	0.481	.000	1	0.991	1.005	0.391	2.583
	Constant	21.541	4.01×10^5	.000	1	1.000	2.26×10^{10}		

a. Variable(s) entered on step 1: Malay, Chinese, Indian, Sabah, Gender, Age, Region, CGPA, Course, Intern.

From the values in Table 5, it can model:

$$\log\left(\frac{p}{1-p}\right) = \sum_{1}^{n} b_i x_i + c \tag{8}$$

$$\log\left(\frac{p}{1-p}\right) = 1.942x_4 - 1.585x_5 - 21.541 \tag{9}$$

Where *p* is the employed status, x_4 = region (urban) and x_7 = CGPA.

Therefore, the employability of UTM Mathematics Alumni (2016 – 2020) is

$$\text{Employed}, p = \frac{1}{1+\exp\,(1.942x_4 - 1.585x_5 - 21.54)} \tag{10}$$

4.2.2 Furthering Study Prediction Analysis

From Table 6, the univariate analysis shows that for confidence interval α = 0.05, the predictor age, x_3, region, x_4, CGPA, x_5 and course program, x_6 has a significant *p*-value. Age, x_3 predictor is a continuous variable. The exp(*B*) value shows that older alumni are 0.554 at odds to younger graduates to further their studies. From the log of odds value, we can conclude that younger graduates have 0.591 more tendencies to further their studies than older alumni. For region, x_4 predictor, exp(B) is graduates who live in the Urban zone are 0.287 at odds with their Rural counterparts in getting employed. Therefore, graduates who live in the Rural region are 1.248 times more likely to pursue their studies than their Urban peers. CGPA, x_5 predictor is also a continuous variable. The exp(B) value shows that graduates with a higher CGPA are 8.859, at odds to their lower CGPA peers in furthering their studies. Taking the log of odds, we can conclude that higher CGPA alumni have 2.181 more tendencies to further their studies than their peers who have lower CGPA. Lastly, for the course program, x_7 alumni who took BSc Industrial Mathematics are coded as '1' and BSc Mathematics is '0'. From the exp(B) value, graduates who took BSc Industrial Mathematics are 0.325 at odd to those who took BSc Mathematics. So, the log of odds are, BSc Mathematics graduates are more likely at 1.124 to pursue Master's Degree and PhD after graduating than their BSc Industrial Mathematics peers.

Table 6: Univariate Analysis in Furthering Study

Variables	Wald	df	*p*-value	Exp(*B*)
Demographic Background				
Malay	0.339	1	0.561	1.481
Chinese	0.031	1	0.860	1.158
Indian	0.007	1	0.933	0.910
Bumiputera Sabah	0.000	1	0.999	0.000
Bumiputera Sarawak	0.000	1	1.000	0.000
Gender	0.262	1	0.609	0.726
Age	6.568	1	0.010*	0.554
Region	6.014	1	0.014*	0.287
Education Background				
CGPA	5.942	1	0.015*	8.859
Course Program	4.598	1	0.032*	0.325

Note: * $p < 0.05$

4.2.2 Underemployability Prediction Analysis

Based on Table 7, for alumni who experience no income change factors, the predictors age and job experience have significant *p*-value. For age, older graduates are at odds 0.664 to younger graduates. Hence, younger mathematics alumni are 0.409 likely to have no income change than the older alumni. Meanwhile, for Job Experience, experienced alumni are at 0.366 odds to entry-level alumni. Therefore, entry-level mathematics graduates are 1.005 more likely to have income no change than the experienced alumni.

Whereas for more workload prediction, the predictors Malay, female and job experience have significant *p*-value. Malay graduates are at odds 0.664 to non-Malays peers. Hence, Malay mathematics alumni are 1.466 more likely to have more workload than the non-Malay alumni. For gender, female alumni are at odds 11.00 to male counterparts. So, female Mathematics have 2.380 more tendencies to get more workload in employment than their male counterparts. Meanwhile, for Job Experience, experienced alumni are at 2.685 at odds with entry-level alumni. Therefore, experienced level mathematics graduates are 0.988 more likely to have more workload from their respective jobs than the entry-level alumni.

Table 7: Univariate underemployment analysis

		No Income Change		***More Workload***	
Variables	**df**	***p*-value**	**Exp(*B*)**	***p*-value**	**Exp(*B*)**
Demographic Background					
Malay	1	0.391	0.615	0.032*	4.333
Chinese	1	0.096	3.611	0.092	0.159
Indian	1	0.266	2.000	0.706	0.625
Bumiputera Sabah	1	0.999	0.000	0.445	0.407
Bumiputera Sarawak	1	1.000	0.000	1.000	0.000
Gender	1	0.070	0.314	0.025*	11.000
Age	1	0.031*	0.644	0.070	1.406
Region	1	0.014	0.287	0.032*	4.333
Education Background					
CGPA	1	0.690	1.287	0.973	0.973
Course Program	1	0.198	0.671	2.098	2.098
Employment Background					
Self-Employed	1	0.203	2.773	0.132	0.190
Job Experience	1	0.038*	0.366	0.047*	2.685
Income	1	0.067	0.492	0.123	1.740
Work Place	1	0.756	1.179	0.730	1.190
Organisation Type	1	0.265	0.292	0.471	1.778
Business Type	1	0.853	0.912	0.995	1.003
Job Sector	1	0.749	0.791	0.687	0.762

Note: * $p < 0.05$

5 Conclusion

5.1 Summary

UTM Mathematics Alumni employability and underemployment can be better visualised and understood. The analysis reveals that mathematics graduates have high employability and stable job. The employability rate of UTM Mathematics graduates is 94.02%. The study also proves that COVID-19 only affect their employment at a minimal rate.

From the analysis, it is found that CGPA does not determine one's employability. It also applied to other categorial factors. No factor correlates to employability.

From logistic regression, it shows that graduates who live in urban regions and have internships related to their studies are more likely to be employed. Therefore, race and gender do not play the central factor in determining Mathematics graduates' employability. The same goes for their income change. Only job experience and age determine their income to have no change. However, in underemployment prediction, Malay and female workers are more likely to experience more workload than their non-Malay and male peers.

5.1 Suggestions

Modified and advanced logistic regression modelling would better fit this data set since all three methods used for ordinary logistic regression all scored less than 50% goodness of fit to this dataset. The alumni information should always keep and get updates annually of their whereabouts. The information gathered would give us better insights into the current employment state for Mathematics graduates. This research only shows the preliminary analysis of Mathematics graduates employability and underemployment post-COVID-19 in Malaysia. In the following years, it is recommended for Mathematics' employability and underemployment research to be done annually.

References

[1] Rolling Updates on Coronavirus Disease (COVID - 19). (2020, July 31). Retrieved from World Health Organization: http://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen

[2] Employment Information and Analysis Services (EIAS), Employment Insurance System Office. (2020). Interim Report 2020 Employment Insurance System Office. Putrajaya: PERKESO. Retrieved from https://eiscentre.perkeso.gov.my/reports-articles/#dearflip-df_8969/1/

[3] Hilbe, J. M. (2015). *Practical Guide to Logistic Regression.* Arizona: CPC Press

[4] Roslly, M. N., & Mohd Khalid, Z. (2020). Modelling Occupations Suitability Amongst Graduates Using Logistic Regression. 34-42.

[5] El-Temtamy, O., & O'Neill, K. K. (2016). Undergraduate employability training and employment: a UAE study. Higher Education, Skills and Work-Based Learning, 6(1), 100 - 115. doi: 10.1108/HESWBL-02-2015-0006

[6] Qian, Y., & Fan, W. (2020). Who loses income during the COVID-19 outbreak? Evidence from China. Research in Social Stratification and Mobility, 68, 1 - 5. Retrieved from https://doi.org/10.1016/j.rssm.2020.100522

[7] Umaña-Hermosilla, B., de la Fuente-Mella, H., Elórtegui-Gómez, C., & Fonseca-Fuentes, M. (2020). Multinomial Logistic Regression to Estimate and Predict the Perceptions of Individuals and Companies in the Face of the COVID-19 Pandemic in the Ñuble Region, Chile. *Sustainability*, 1-20. doi:10.3390/su12229553

[8] Harrell Jr, F. E. (2015). Regression Modeling Strategies With Application to Linear Models, Logistic and Ordinal Regression and Survival Analysis. Nashville, Tennesse, United States of America: Springer. doi: 10.1007/978-3-319-19425-7

[9] Glen, S. (2016, August 28). Hosmer-Lemmeshow Test Definition. Retrieved from StatisticianHowTo.com: https://www.statisticshowto.com/hosmer-lemeshow-test/