

EXTENDED ABSTRACT

PREDICTING UNDERGRADUATES' EMPLOYABILITY USING NAÏVE BAYES CLASSIFICATION METHOD

Premathilaka W.S.C.S^{*} and Imalka K.H.J

University of Ruhuna, Sri Lanka

* chaminipremathilaka@gmail.com;jayani@ictec.ruh.ac.lk

(Published 15 October 2021)

Abstract

Graduate employability is a crucial concern to all countries in the world. Many graduates are looking for job opportunities despite their qualifications and skills due to various factors. Therefore, predicting the employability of undergraduates considering their performances has become a timely requirement as it can help to address unemployment in great success since it provides students an opportunity to groom. This research analyses demographic and educational factors affecting an individual's employability and develops a model to predict an undergraduates' capability of achieving career targets beforehand. Data mining methods are often implemented nowadays for analyzing available data and extracting information and knowledge to support decision-making in various domains. In this study, the Naive Bayes classification algorithm is used to predict the possibility of undergraduates' employability. This paper presents the results of potential employability of undergraduates using demographic and educational information collected via an online survey conducted through social media platforms. Apart from demographic factors, undergraduates' lecture participation rate, computer literacy, English literacy, intern participation, club memberships, and many more attributes were used to evaluate their competencies. Overall, individuals actively involved with academic activities, including extra-curricular activities, were more likely to be predicted as employed. Therefore, using the naïve Bayes classification algorithm, the most influencing factors that can be used to predict undergraduates' employability are presented by using a trained model with graduate data with an accuracy of 75.89 %.

Keywords: Data mining classification, Naive Bayer's, graduate unemployment

1. Introduction

Unemployment describes people who are employable and actively seeking employment but unable to find a job. Over the previous few decades, graduate unemployment has been a significant issue in Sri Lanka (Ariyawansa, 2008). It is claimed that young people want to find their place in society. Universities are in charge of answering some crucial concerns about their mission and future strategy, as well as their alumni's current standing. This research is done by analyzing and comparing the extracurricular activities and other activities which the current employed graduates and unemployed graduates have done during their university life in order to achieve qualifications, such as leadership roles played, sports done, lecture attendance rate, gender, place of origin (urban/rural), CGPA¹

©University of Vavuniya, Sri Lanka 2021.

1. CGPA is the academic performance from 0.00–4.00, calculated as an average of the GPA from all completed terms/semesters.

and English literacy. The parameters indicated above were compared to current undergraduate data, using data mining techniques. The likelihood of employability or unemployability of the undergraduates was predicted using those data. Undergraduates will be able to develop their abilities and external qualifications, which are essential to obtain successful work in the future, using the knowledge and results gained via this research. A survey was conducted to gather information about currently employed graduates, unemployed graduates, current undergraduates, and details on what graduates did throughout their university years and what undergrads are presently doing. Universities are now pursuing changes to their teaching-learning programs to meet the current demands of a demanding business environment (Singam, 2017). With current statistics indicating that the unemployment rate is rising, it is critical to investigate the skills gaps students must fill.

2. Methodology

The data set for this study contains demographic data from both graduates and undergraduates. The data set was separated into two samples as graduate dataset and the undergraduate dataset. The final dataset utilized for research implementation has 21 nominal and numeric attributes. University, Faculty, Gender, Hometown, “Are you employed” (Employability), GPA, Class, Other External Courses, Lecture Participation rate, English Literacy, Computer Literacy, and extracurricular activities are some of the variables connected with graduate and undergraduate statistics. The graduate dataset consists of 31.06%(96) samples, and the undergraduate dataset consists of 68.9%(213).

A classification algorithm was used to predict undergraduate employability. Since the classification model attempts to draw conclusions based on the input values provided for the training data set (graduate data), this method predicts the class categories for new data (undergraduate data). In this study, the Naive Bayes classification algorithm is used as this requires a small amount of training data to estimate the necessary parameters and is extremely fast compared to other classification methods. Figure 1 depicts the entire model developed to predict undergraduate employability. The model was initially trained and tested with 96 graduate samples. Retrieve operator was used to retrieving graduate data.

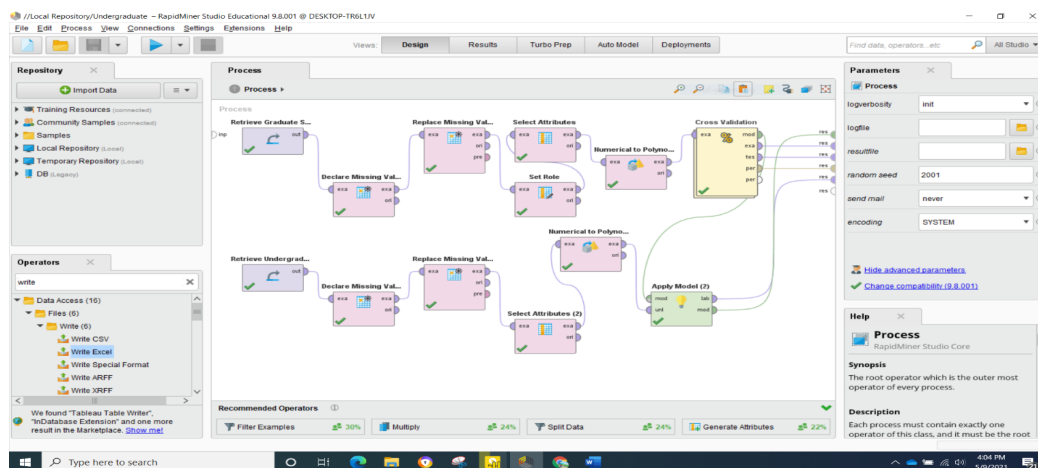


Figure 1. Employment Model

Declare missing value and the replace missing value attributes were used to pre-process the data. For the set role attribute variety of labels, including class, clubs, and computer literacy were use and as the prediction attribute “Are you Employed” attribute was used, as illustrated in Figure 2. Select attribute operator shown in Figure 1 is used to choose a subset of the attributes of ExampleSet and to remove the other attributes. Numerical to Poly-nominal operator is used in the model to convert the numerical values into poly nominal type.

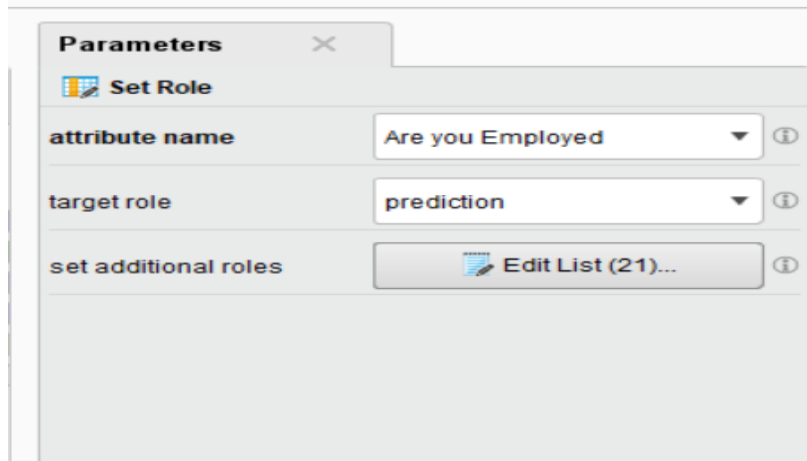


Figure 2. Prediction parameter

The cross validation operator is used in the model to estimate the statistical performance of the learning model. The model was trained using the Naive Bayes operator in the Training subprocess. As illustrated in fig. 3, the trained model was then used in the Testing subprocess, with the performance classifier operator serving as the testing subprocess of graduate dataset. The Bayes Operator was used to train the model on the Graduate’s dataset. Naive Bayes is a low-variance, high-bias classifier that can develop an effective model even with a tiny dataset (Garg, 2018). The Naive Bayes method is simple to use and has a low computing cost.

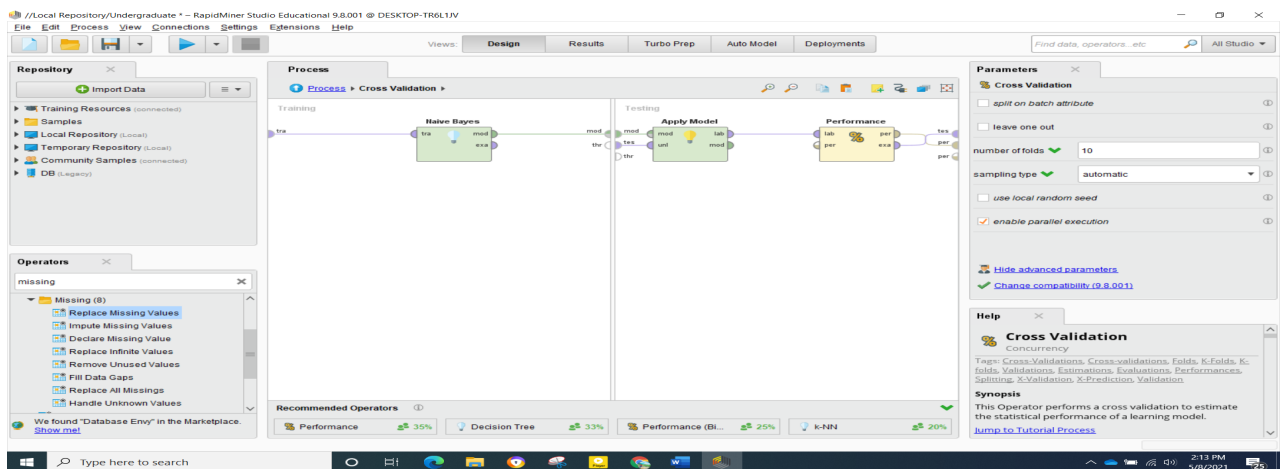


Figure 3. Training and Testing subprocess of cross validation operator

For statistical performance evaluation of classification tasks, the Performance (Classification) operator was implemented. After the model has been trained using the graduate dataset and has achieved the required output, the undergraduate dataset is connected to the model to predict undergraduate employment. Finally, the model’s projections are used to determine undergraduate employability.

3. Results and Discussions

The model was trained using 96 graduate sample data, and could obtain 75.89% of total accuracy, 0.531 correlation and 1.214 cross-entropy. As shown in Figure 4 in total we had total 32+10 “Are you employed - No” candidates and when we the algorithm is executed, it was predicted that “Are you employed - No” 32 times, which were actual “True - Are you employed - No “and 10 times it

predicted “Are you employed – yes”, where in fact it should have been “Are you employed – no”. In total we were able to capture 76.19%, “Are you employed – No” cases. In the same way for the “Are you employed – Yes” cases, we have achieved 75.93%.

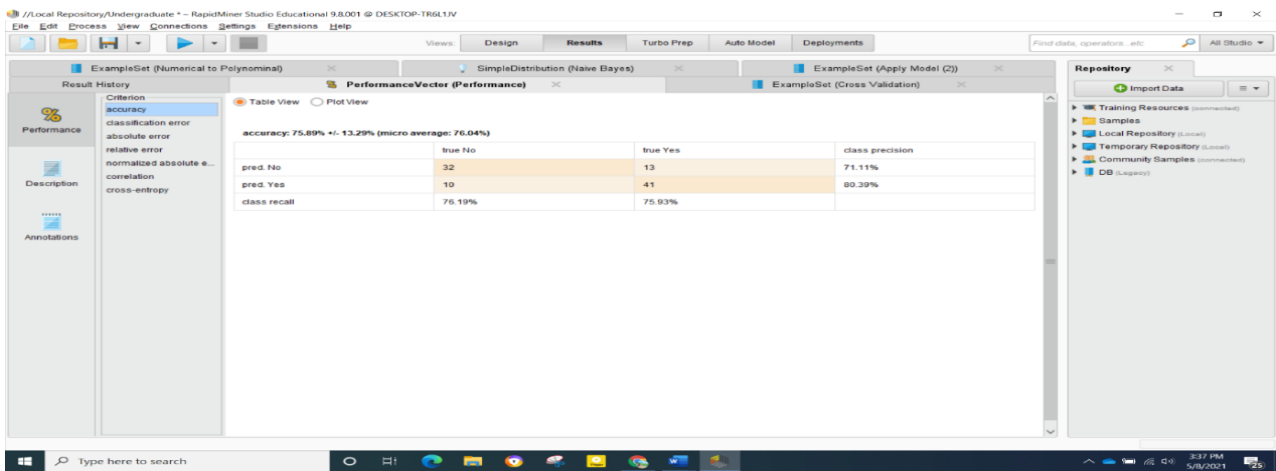


Figure 4. Employability Prediction Accuracy

In the model, the graduate dataset is used to train the model and then the trained model is applied to a set of 213 undergraduate data to predict the employability possibility of undergraduates in future. Individuals who had done more extra work beyond their degree had a better probability of getting hired than those who had done less extra work. The percentage of students who attend lectures, English literacy, and computer literacy all have a role in graduates’ career prospects. Undergraduates who had done more extra work were expected to have a better chance of being hired in the research. In order to achieve a successful employment opportunity, it is essential to develop one’s talents and polish one’s abilities. Attending university conferences, volunteering, and performing community service, joining university organizations, completing an internship, and other extracurricular activities can assist students in enhancing their talents while also strengthening their interpersonal skills.

The study’s primary limitations were the limited dataset and the inability to acquire a comprehensive dataset due to the continuous pandemic scenario from the beginning to the end of the study. Despite utilizing a smaller dataset, we were able to obtain a 75.89 percent accuracy in the research. In the future, a massive dataset from graduates and undergraduates from all universities will be collected to enhance the model’s accuracy and forecast the specific occupation an undergraduate can find after completing their studies.

3.1 Conclusion

Graduate unemployment is a political, economic, and social dialogue that is ongoing. There are countless researches that analyse unemployment, but there are no studies that forecast the probability of employment. The Naive Bayes classification approach was employed in the study to predict employability. After evaluating many hypotheses, some of the most important characteristics were identified and used to train a graduate model to predict undergraduate employment. With this forecast, undergraduates will be able to improve their performance depending on the outcome while still in university.

Students can focus on improving their capabilities by undertaking extracurricular work while studying for their degree. In the future, we want to improve the accuracy of the original model by utilizing a large data set that spans 10 Sri Lankan universities in order to achieve better results, and it is intended to forecast undergraduates’ specific employment.

References

- Ariyawansa, R. G. (2008). Employability of Graduates of Sri Lankan Universities. *Sri Lankan Journal of Human Resource Management*, 14.
- Garg, R. (2018). Student's Performance Evaluation of an Institute Using Various Classification Algorithms. *Singapore: Springer Nature Singapore Pte Ltd.*
- Singam, K. (2017). Review on Graduates' Unemployment in Sri Lanka and the Globe. *Global Journal of HUMAN-SOCIAL SCIENCE: G Linguistics Education*, 8.