Predicting Students At-Risk Using Deep Learning Neural Network

Student: Tshepang Motaung (1431795) Supervisor: Dr. Ritesh Ajoodha

School of Computer Science and Computational Mathematics
University of the Witwatersrand, Johannesburg
South Africa

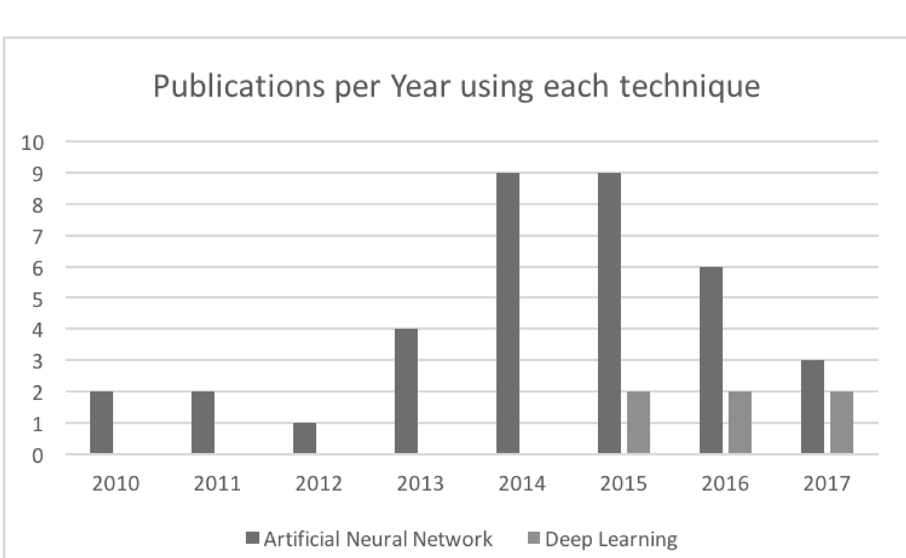
Objectives

The impact of students' interaction with the e-learning system will be examined in this article. Additionally, the data collected will aid schools in enhancing student performance. Assisting administrators in developing learning systems is also a goal:

- To predict at-risk students
- and find factors that mostly affect students' academic performance whether be demographic factor, academic background of students or behavioural factors

Introduction

Students' academic performance is difficult to gauge since it is influenced by a variety of factors such as demographics, educational experience, and other external pressures. Machine learning algorithms have lately gained popularity for analyzing educational data and uncovering hidden meaningful patterns for predicting student grades. Deep learning allows the model to learn from examples, patterns, or events, rather of having to manually construct the features [1]. Despite the fact that a variety of models have been investigated in the context of learning analytics research, identifying the utility of deep learning in learning analytics is still in its early stages, and research into its implementation has only recently begun. Figure 1 presents a number of authors per year that have worked on EDM using either ANN or DL [1].



Related Work

Several authors' approaches to tackling our problem statement were examined in this section. We investigated the datasets they used, the probabilistic models they used, and other data preprocessing techniques they employed. Several authors took different methodologies, which resulted in them reaching different conclusions. We will present the table of accuracy's and models obtained by other author's:

Table 1:Accuracies Obtained By Other Researchers.

Authors	Model	Accuracy
Amrieh [2015] preprocessing	ANN	73.8%
Waheed [2020] predicting	SVM	85.65%
Mathye [2020]	RF	73%
Huang [2013] predicting	MLP	88.2%
Akour [2020] effectiveness	DT	65%

Method

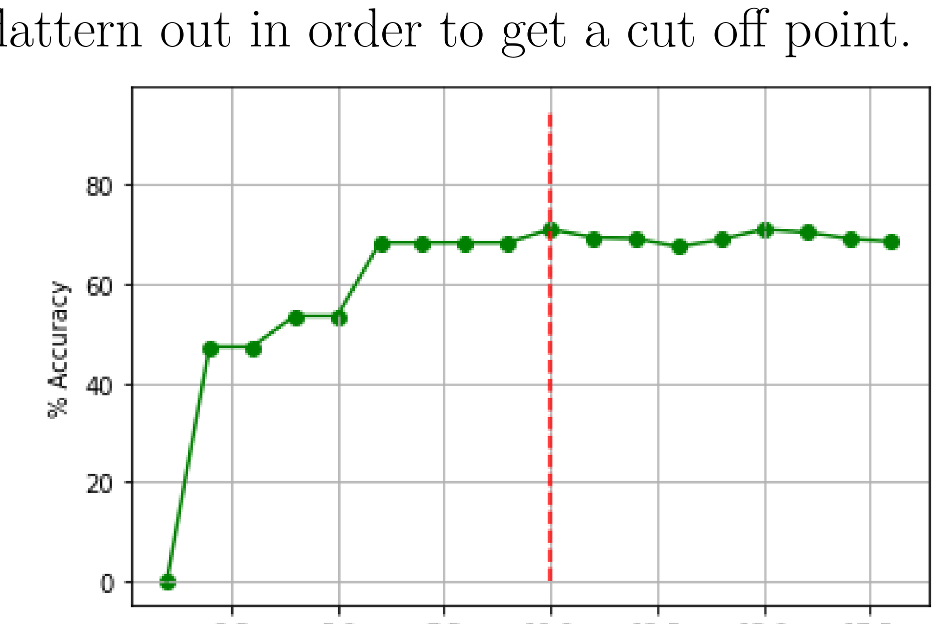
- Our dataset consisted of both numerical and categorical values.
- We utilized One-Hot-Encoding to convert categorical data into numerical inputs. Because our categorical variables have no natural ordering, One-Hot-Encoding was chosen over Label-Encoding.
- Using feature selection, we may remove unnecessary and redundant data, reducing time complexity, improving learning accuracy, and making the learning model and data more understandable.
- To keep and eliminate attributes, information gain ranking was employed.

Important Result

Features used were ParentAnsweringSurvey$_{No}$, $ParentAnsweringSurvey_{Yes}$, $Parent_{Father}$, Parent$_{Mom}$, $AbsenceDays{>}7$, $AbsenceDays{<}7$, $ParentSchoolSatisfaction_{Good}$, ParentSchoolSatisfaction$_{Bad}$, $Gender_M$, $Gender_F$, VisitedResources, RaisedHands, AnnouncementView and Discussion.

Feature Selection

Figure 1 presents a method of feature selection when we continuously add features with accordance to their mutual gain until the model performance starts to flattern out in order to get a cut off point.



Results

The experimental results of the classification utilizing the pruned feature set is presented by Table 2 below:

Classifier	Accuracy	Time to build the model
kNN	0.7574	0.075 sec
RF	0.7736	0.084 sec
SVM	0.7751	0.010 sec
MLP	0.7373	0.088 sec
CNN	0.7258	2.00 sec

Table 2:Models Classification Accuracy With Pruned Features

Conclusion

The classification of student academic performance was achieved in this study by first identifying and extracting elements that affect student performance. In order to make the feature vector simple, thorough, and efficient, students' features were examined for their significance in the task of performance categorization. SVM, which outperformed the deep learning models MLP and CNN, was found to be the most accurate model after experimenting with a range of off-the-shelf classifiers, with an accuracy of 77.51%. We discovered that behavioral elements have a favorable impact on students' academic progress using data preprocessing approaches.

Additional Information

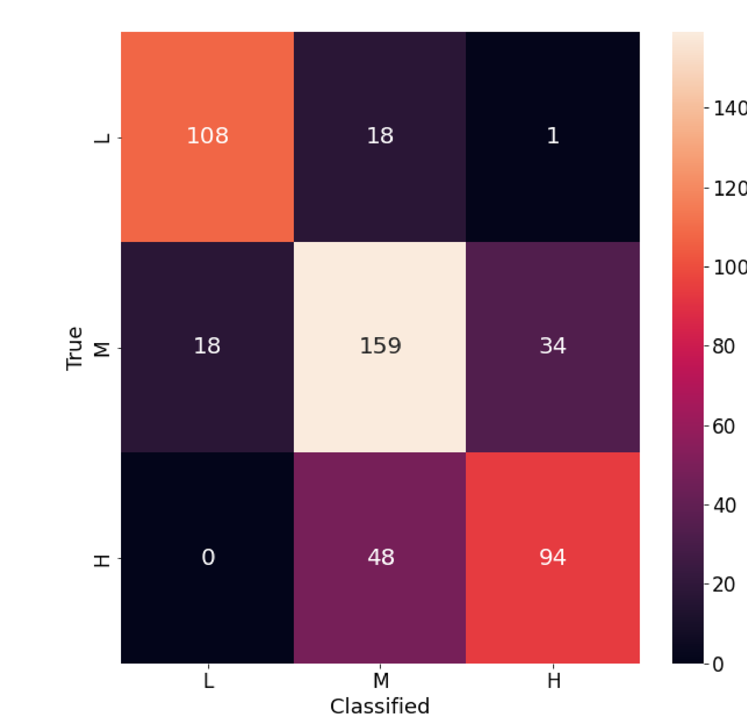


Figure 2:Confusion Matrix Of The Best Performed Model (SVM)

References

[1] Orlando Bisacchi Coelho and Ismar Silveira. Deep learning applied to learning analytics and educational data mining: A systematic literature review. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 28, page 143, 2017.

[2] Elaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah. Preprocessing and analyzing educational data set using x-api for improving student's performance. In *2015 IEEE Jordan Conference on Applied Electrical*