

# WHAT'S NEXT? Prediction Model for Students Future Development

Neethu C Sekhar<sup>1</sup>, Merlin Sebastian<sup>2</sup>, Nandita Suresh<sup>3</sup>, Leoyon Reji<sup>4</sup>, Shahad C K<sup>5</sup>

<sup>1</sup>Assistant Professor, Department of Computer Science and Engineering, Amal Jyothi College of Engineering (APJ Abdul Kalam Technological Kalam Technological University) Kottayam, India

<sup>2,3,4,5</sup>Department of Computer Science and Engineering, Amal Jyothi College of Engineering (APJ Abdul Kalam Technological Kalam Technological University) Kottayam, India

**Abstract** - Categorizing and predicting the future of students based on their exam performance are typical challenges for educators. Therefore, proper and accurate method is very important to give timely advice to students regarding choice of program and university. Previously, traditional data mining methods, such as decision trees and association rules, were used for classification. The rapid development of artificial intelligence and deep learning algorithms in recent years offers a different method for intelligent classification and prediction of results. TensorFlow's artificial intelligence engine helps to classify students' performance and predict their future university. In our system, the deep learning model analyzes subjects like physics, mathematics, English, chem, bio, history and Malayalam, as well as nonacademic achievements like driving, sports, and art. With a dataset of 2500 students, 80 per cent of that data is used as training data and 20 per cent is used as test data, accuracy ranges from 97 per cent to 99 per cent. The optimal configuration of the TensorFlow deep learning model to achieve the best prediction accuracy is determined.

**Index Terms** – Artificial Intelligence, Deep Learning, future prediction, Convolution Neural Network, Recurrent Neural Network, TensorFlow, prediction modelling.

## I. INTRODUCTION

How to predict student's learning output has always been a concern for students, teachers and parents. On the basis of the results of past examinations and class evaluations, you can predict the future development of your students. This is a difficult problem as it contains a large amount of data in the education database and the results can affect the future development of children. Accurate and good predictions can benefit and impact students, educators and institutions. Over the decades, different types of

data mining techniques have been used for performance prediction (eg: decision trees, Naive Bayes, K Nearest Neighbors and Support Vectors). However, for the development of artificial intelligence and deep learning applications, it is becoming more important to use AI tools like Google Tensorflow for pattern recognition today. In this project, we will study how to use artificial intelligence and deep learning algorithms for the correlation between pattern recognition and evaluation results.

## II. LITERATURE SURVEY

Some of the strategies referenced in the literature are briefly summarised here.

In[1], Predicting a student's overall performance turns into greater tough because of the big quantity of information in academic databases. Currently in Malaysia, the lack of a current machine to investigate and display student development and overall performance isn't being addressed. There are many reasons for this. First, research on current forecasting strategies is still insufficient to find the most suitable strategy for forecasting the overall performance of Malaysian university students. The second reason is the lack of research on the factors affecting student performance, especially publications within Malaysia. In [1] therefore, a systematical literature assessment on predicting student overall performance through the usage of information mining strategies is proposed to enhance college students achievements. The major goal of this paper is to offer an outline of the information mining strategies which have been used to expect college students overall performance.

In[2], Application of classification model in selecting college students who may fail in the first year of study. The data is collected from three groups of college students in the teaching years from 2010 to 2012. Students with inside the institution had been sampled from plenty of disciplines ( $n = 1074$ ) and sundry in phrases of education and educational ability. The signs used to consist of information usually to be had for colleges, which includes age, gender, and beyond instructional performance. [3]They analyzed other validated psychometric indicators that can be evaluated in the early stages after enrollment, especially personality, motivation, and learning strategies.

In the method of [4], the subject is a neural network student performance prediction model. A neural network student performance prediction model (nnsppm) in this discipline. Early prediction of student performance is important for academics to plan strategic interventions before biology students enter the upper grades. This article describes a study on the development of artificial neural network (ANN) models to predict the learning outcomes of engineering students. Use the cumulative grade point average (GPA) of 4,444 to measure academic performance for the eighth semester. The research was carried out at the Department of Electrical Engineering, Teknologi MARA University (UiTM), Malaysia. Student performance in core subjects in the first semester is used as an independent or predictor variable, and the ACMG in the eighth semester is used as an output or dependent variable. The work of [5], the investigation was carried out for two different entry points, registration and diploma. The performance of the model is measured by the correlation coefficient  $R$  and the mean error (MSE). Research results show that first and third-semester core courses have a large influence on final GPA upon graduation.

### III. METHODOLOGY

Real-world data generally contains noise, missing values, and maybe in unusable formats that cannot be used directly in machine learning models. It is necessary to preprocess the data to clean the data and make it fit the machine learning model, which also improves the accuracy and efficiency of the machine learning model. It involves the following steps:

**Obtaining a data set:** To create a machine learning model, we first need a data set, because the machine learning model works entirely with data. Data collected in an appropriate format for a particular problem is called a dataset. To use the dataset in code, we usually put it in a CSV file.

**Importing libraries:** To run data preparation with Python, you need to import some predefined Python libraries. This library is used for a variety of tasks.

**Importing datasets:** We must now bring the collected dataset for the machine learning project. However, you must set the working directory to the current directory before importing the dataset. We can set the directory in the working directory, but you must include a dataset you need. Now, to import a dataset, use the playback CSV file used to play the CSV file and perform a variety of tasks. You can use this feature to play the CSV file locally via the URL.

**Encoding Categorical Data:** Preprocessing of machine learning data splits the data set into a training set and a test set. This is one of the critical steps in data preprocessing. Because this can improve the performance of your machine learning model. Say you trained a machine learning model on one dataset and tested it on a completely different dataset. This makes it difficult for the model to understand the correlation between models.

**Splitting dataset into training and test set:** Divide the data set in the preprocessing of machine learning data into a training set and a test set. This is an important step in data preprocessing to improve the performance of machine learning models. Suppose you train a machine learning model as one dataset and test it on another dataset.

**Feature scaling:** The scaling of the function is the last step of the pretreatment of the data of the learning of the machine. A technique for standardizing the independent variables of a dataset within a specific range. In functional scaling, we will leave the variable with the same range as the same ladder and does not dominate other variables. Deep learning is a class of machine learning algorithms that use multiple layers to extract high-level functionality with native input. CNN consists of two basic parts of the extraction of

functionality and classification. Function extraction includes several convolutional layers, maximum pooling and activation. The classifier is usually composed of fully connected layers. One of the main parts of the neuronal network is a neuronal convolution network (CNN). They consist of neurons with learning weights and deflections. Each concrete neuron receives a lot of entries and passes through the activation capabilities and responds again to the output.

Dataset: Since datasets are not available from any sites we created a new dataset for this project. These 2500 datasets are divided into 2. The first dataset containing 80 percentage of the dataset, i.e. 2000 datasets, was defined as training data, while the other 20 percentage, i.e. 500 datasets, were defined as training data. used as test data. Assessment results from different subjects are used as input  $x_1, x_2, \dots, x_n$ . The result of each program and university selection is used in the outputs  $y_1, y_2, \dots, y_n$ .

#### IV. PROPOSED SYSTEM

Neural networks are a new technology for education and data mining. The advantage of neural networks is that they can detect all possible interactions between predictors. With more computing power available today, more neural network hierarchies can be implemented and deep running analysis can be implemented. Deep learning can also detect complex non-linear relationships between dependent and independent variables. This project uses the Google Tensorflow Deep Learning Analysis Engine to predict the future development of students. The attributes analyzed are traditional subjects, including academic achievement and behavior such as mathematics, Malayalam, English, physics, chemistry, biology, history, and non-academic scores such as sports, arts and participation. we are analyzing the service and 2,500 records will be generated based on specific rules. For example, students who are good at physics and mathematics can take engineering courses. Initially, we split these 2500 datasets into two parts. The first data section contains 80 percentage of the dataset set in the training data, or 2000 datasets, and uses the remaining 20 percentage of the data, or 500 datasets, as test data. increase. The performance of each subject evaluation is used as input for  $x_1, x_2, \dots$

$x_n$ . The result of the subject and its university selection is used for the outputs  $y_1, y_2, \dots, y_m$ . A CNN (Convolutional Neural Network) model is used to build a predictive model. CNN's are often used for pattern recognition problems. Using one hot output system, a fully connected network is set up and the relationship with another node is  $Wx + b = y$ . The softmax cross-entropy is used to evaluate the cost function, and the main structure is completed as shown in the following figure.

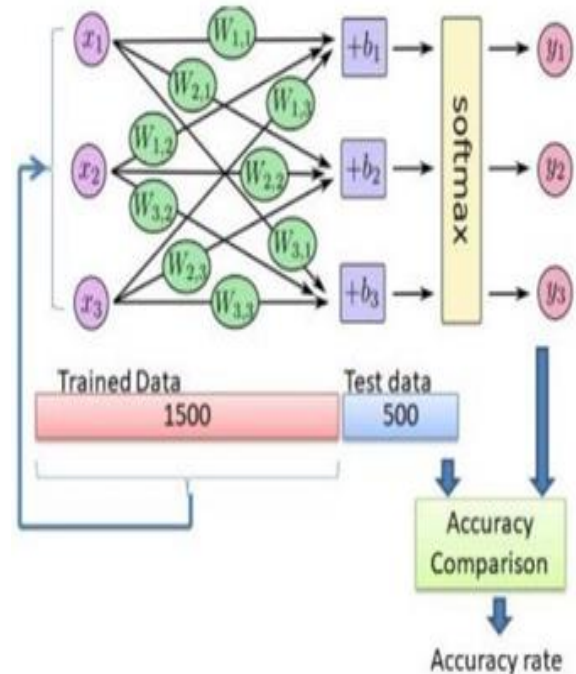


Fig. 1. CNN Architecture

This project uses Python and Tensorflow deep learning engines for development. It is a symbolic math library that is also used in machine learning applications. For example, TensorFlow is an open-source software library for multitasking data flow programming. Neural Networks. Deep learning in the cloud makes this survey easy to start. This framework has been widely recognized by the industry and has become a popular choice for deep learning research and application development, especially in the fields of computer vision, natural language understanding, and speech translation. Student performance and future development follow certain rules, so the pattern recognition engine is used for analysis. After the model was trained, we tested the performance of the trained model.

System Architecture

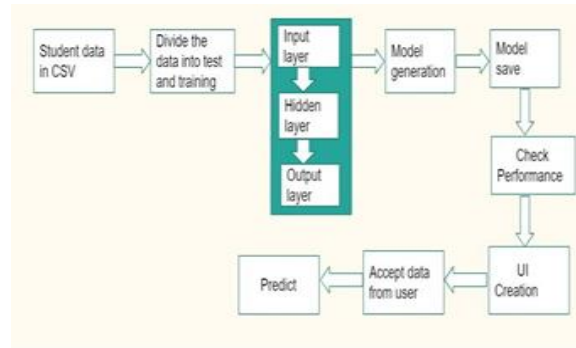


Fig. 2. System Architecture

## V. EXPERIMENTAL RESULT AND ANALYSIS

The student prediction model is created using the CNN model and using it, we obtained an accuracy rate of 97 percentage to 99 percentage. Fig3 shows the accuracy of the cnn model with 10 epochs.

we then compare this CNN model with the rnn model, which has an accuracy of 40 percentage to 50 percentage. Fig4 shows the accuracy of rnn model with 100 epoch out of which accuracy obtained in last 5 epoch is given in the image.

we can see that the CNN model is more efficient for creating student prediction models than the rnn model. Now we check the performance of the model created by getting data from the dataset and predicting its output. The model prediction is correct, and then the user provides random input data and predicts the output, and the model also predicts correctly.

You can see that the deeper the deep learning model and the more efficient, the more accurate the results are and not always accurate. There are sweet spots that must be tested and identified. For learning rate, The higher the rate, the faster the learning model converges. But if the rate is too high, the result can exceed the sweet spot. Therefore, it is possible to improve the prediction performance for low momentum and high rate. After that, you can gradually increase the momentum and reduce the rate to guarantee convergence. This study shows that deep learning can be an effective tool for predicting student outcomes. Results range from 97 percentage to 99 percentage. The predicted results are sufficient to provide appropriate suggestions so that students, teachers and parents can determine their developmental pathways. I think more deep learning applications will be used to educate and train corporate employees.

```

Epoch 1/10
16/16 [=====] loss: 0.1263 - accuracy: 0.9760
Epoch 2/10
16/16 [=====] loss: 0.0958 - accuracy: 0.9880
Epoch 3/10
16/16 [=====] loss: 0.0970 - accuracy: 0.9793
Epoch 4/10
16/16 [=====] loss: 0.0957 - accuracy: 0.9700
Epoch 5/10
16/16 [=====] loss: 0.0701 - accuracy: 0.9887
Epoch 6/10
16/16 [=====] loss: 0.0551 - accuracy: 0.9920
Epoch 7/10
16/16 [=====] loss: 0.0553 - accuracy: 0.9907
Epoch 8/10
16/16 [=====] loss: 0.0478 - accuracy: 0.9907
Epoch 9/10
16/16 [=====] loss: 0.0420 - accuracy: 0.9933
Epoch 10/10
16/16 [=====] loss: 0.0423 - accuracy: 0.9940
  
```

Fig. 3. Accuracy of CNN model

```

Epoch 90/100 loss: 1.5541 - accuracy: 0.2838
16/16 [=====]
Epoch 91/100 loss: 1.5544 - accuracy: 0.2885
16/16 [=====]
Epoch 92/100 loss: 1.5530 - accuracy: 0.2818
16/16 [=====]
Epoch 93/100 loss: 1.5553 - accuracy: 0.2818
16/16 [=====]
Epoch 94/100 loss: 1.5561 - accuracy: 0.2758
16/16 [=====]
Epoch 95/100 loss: 1.5563 - accuracy: 0.2851
16/16 [=====]
Epoch 96/100 loss: 1.5566 - accuracy: 0.2851
16/16 [=====]
Epoch 97/100 loss: 1.5575 - accuracy: 0.2851
16/16 [=====]
Epoch 98/100 loss: 1.5576 - accuracy: 0.2851
16/16 [=====]
Epoch 99/100 loss: 1.5579 - accuracy: 0.2851
16/16 [=====]
Epoch 100/100 loss: 1.5558 - accuracy: 0.2851
16/16 [=====]
  
```

Fig. 4. Accuracy RNN model

## VI. CONCLUSION

Classification and prediction are common issues. Existing data mining techniques such as association rules, decision trees, and clustering have been used for decades to solve this problem. As deep learning and artificial intelligence using Tensorflow have grown in popularity, new methods and directions have been opened up to solve the problems of classification and prediction of non-linear results. In this study, the number of hidden layers, hidden nodes, number of

repetitions and hackspru'l were adjusted and compared.

#### REFERENCES

- [1] Amirah Mohamed, Shahiri Wahidah, "A Review on Predicting Student's Performance Using Data Mining Techniques", *Procedia Computer Science*, Volume 72, 2015,
- [2] P. M. Arsad, N. Buniyamin, J.-I. A. Manan, A neural network students' performance prediction model (nnsppm), in: *Smart Instrumentation, Measurement and Applications (ICSIMA)*, 2013 IEEE International Conference on, IEEE, 2013,
- [3] Joonhyuck Lee, Dongsik Jang and Sangsung Park, "Deep Learning Based Corporate Performance Prediction Model Considering Technical Capability", 2017.
- [4] Rianne Conijn, Chris Snijders, Ad Kleingeld and Uwe Matzat, "Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS" October 2016 *IEEE Transactions on Learning Technologies*
- [5] Sadiq Hussain, Zahraa Fadhil Muhsion, Yass Khudheir Salal, Paraskevi Theodorou, Fikriye Kurtog'lu, G. C. Hazarika. "Prediction Model on Student Performance based on Internal Assessment using Deep Learning"