

# A Job Recommendation and Prediction with Uncertainty Estimation Using RNN

**R.Sathviki, B.Jeevitha, G.Akshay, D. Saravana**

Department of Computer Science and Engineering, K L Deemed to be University, Vaddeswaram, Guntur, Andhra Pradesh

## Abstract

### Article Info

**Volume 83**

**Page Number: 11069 - 11075**

**Publication Issue:**

**May - June 2020**

Currently, college-going students would take more time over their parental generations. Further, in the united .States, the six-year graduation rate need been 59% for decades. Moving forward those educational quality by preparing better-prepared learners who could effectively propose over An auspicious way may be basic Also too to anticipate those part. Faultlessly foreseeing students' occupation part clinched alongside future need pulled in substantially consideration as it could assistance distinguish good way might make given will them on time by advisors. Former Look into once students' part prediction incorporate shallow straight models; however, students' scholastics and investment will be a Exceptionally mind boggling transform that includes those amassing of information crosswise over an arrangement for parts that could not be sufficiently displayed by these straight models. What's more to that, former methodologies concentrate on prediction precision without recognizing prediction uncertainty, which may be key to advising and choice making. In this work, we avail recurrent neural network (RNN). These c's models need aid In light of those suspicion that former information of the scholar can provide people for future occupation part thus that evaluations from claiming former courses could make used to anticipate evaluations in An future course. The MLP ignores the transient flow from claiming students' information advancement. Hence, we recommend RNN to students' execution prediction. Should assess those execution of the suggested models, we performed broad investigations ahead information gathered. Those test Outcomes indicate that the recommended models accomplish finer execution over former state-of-the-craft methodologies and give more exact outcomes as a result.

**Keywords :** Sequential Models, Educational Data Mining, Job Prediction

### Article History

**Article Received:** 19 November 2019

**Revised:** 27 January 2020

**Accepted:** 24 February 2020

**Publication:** 19 May 2020

## I. INTRODUCTION

In the educational data mining field Academic students' performance is one of the most important topics.

It can be a very useful factor for both students and teachers. Data mining conception basically extract the meaningful data from the raw information and the Educational Data Mining (EDM) is a developing technique which extracts meaningful knowledge in the context of educational field. Researchers

11069

contributed a lot on EDM theory to learn and investigate educational psychology and also data sciences that closely related to the learning analytics.

Many different machine learning approaches are there to forecast a student's presentation using the student's family details as well as educational background. Particularly, these machine learning method enable to detect the student, who is at the risk i.e. who is likely to fail or drop out from the class. This helps a teacher to provide a remedy to the weak student. It can also help to identify the high performer student in the class.

Decision Tree [1] and Naive Bayes [2] algorithm is highly used in Educational Data Mining. HavanAgrawal [3] stated limitation of these algorithms, the accuracy of the Bayesian classification model reduces when input is provided in a continuous range.

Artificial Neural Network is a powerful tool to make a prediction. Deep learning is the state of art [4] for artificial intelligent research. Currently deep learning concept is highly used in voice/sound recognition [5], Natural Language Processing [6], computer vision [7].

Here proposed Recurrent Neural Network (RNN) classifier model to predict the students' Job. RNN is a different of a neural network. It obtains the output value based on the past and the current information using the recurrent loop of the network. The proposed RNN model aims to predict a student Suitable Job through logistic classification analysis. One hidden layer with Relu activation function is implemented. The proposed model is able to predict a student Job with the accuracy of around 85%.

## II. LITERATURE SURVEY:

Kyndt et al.[8] Utilized neural network to predict academic execution of the first year b.tech student, In light of students' motivation, methodologies with

learning, attempting memory limit and consideration.

Ioannis E. Livieris, et al. [9] Here the studytells about the performance of students in Mathematics using an Artificial Neural Network (ANN) classifier. They Discovered that the changed ghastly perquisite prepared simulated neural system performs finer arrangement contrasted with different classifiers in this context.

S. Kotsiantis, et al. [10] investigated in distance learning using machine learning techniques for dropout prediction of students. Their study made an important contribution as it was a pioneer and helped to carve the pathway for educational data mining.

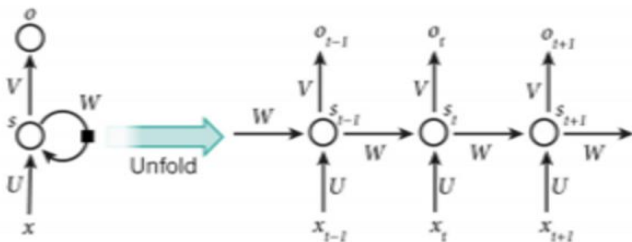
Moucary, et al. [11] utilized anhybrid strategy on K-Means clustering Also simulated neural system to foreseeing those execution about people of higher training about new outside dialect. They utilized neural system to foresee the student's execution et cetera fitting them for An specific cluster utilizing the K-Means calculation. This clustering aided the instructors to distinguish a student's abilities throughout their right on time phases for scholastics.

Mukta Also Usha[12] predicted those academic execution from claiming business school graduates utilizing neural networks And accepted Factual systems they recognized the underlying constructs of a accepted business school educational program Also exhibited its significance for those different components of confirmation transform.

Amrieh, et al. [13] suggested a prediction model for students' execution In view of information mining routines. Those model might have been assessed previously, three diverse classifiers; Naïve Bayesian, simulated neural system Also choice tree. The classifier's execution might have been enhanced by utilizing irregular forest, packing and boosting as a cluster system. Those model attained dependent .

### III. IMPLEMENTATION

A Recurrent Neural Network (RNN) Is An class of artificial neural system the place associations the middle of units form a graph chart along an arrangement. It is mainly used to handle the time series data. RNN may be on settle on utilization of consecutive majority of the data. RNNs are called repetitive a result they perform those same undertaking for each component omj,=mf a grouping.RNN can be described in other way to say that has “memory” to capture the data that is calculated till now. Figure 1 (a) shows a graphical illustration of a structure of an RNN. Figure 1 (b) shows the recitationby the time of working out of an RNN



**Fig 1:** Structure of the Recurrent Neural Network

Those forward pass of the RNN may be Just about same Likewise that from claiming an MLP with single hidden layer but that actuation arrives in those hidden layer from both the present outside information and the hidden layer activations particular case venture once more in time. To the information to hidden units we have:

$$a_h^t = \sum_{i=1}^I W_{ih} X_i^t + \sum_{h=1}^H w_{hh} b_h^{t-1}$$

The output unit has:

$$a_k^t = \sum_{h=1}^H w_{hk} b_h^t$$

The back propagation of the RNN is just the standard back propagation. Those complete arrangement for delta terms could make computed toward beginning In  $t = t$  and recursively applying those The following functions, decrementing  $t$  In each step. Note that  $\delta_j^{T+1} = 0, \forall j$ , since no slip is gained from past the limit of the arrangement.

$$\delta_h^t = \theta^j(a_h^t) \left( \sum_{k=1}^k \delta_k^t w_{hk} + \sum_{h=1}^H \delta_h^{t+1} w_{hh} \right)$$

$$\delta_h^t = \frac{\partial O}{\partial a_j^t}$$

### DATA PROCESSING

It is an educational data set together since a learning management system. The data set is little bit updated by adding two more features with the existing data set. The data set contains record for total 480 students. The modified data set contains total of 34 features as below

|                                             |                                                           |
|---------------------------------------------|-----------------------------------------------------------|
| 1. Academic percentage in Operating Systems | This feature contains percentile in Operating Systems.    |
| 2. Percentage in Algorithms                 | This feature contains percentile in Algorithms.           |
| 3. Percentage in Programming Concepts       | This feature contains percentile in Programming Concepts. |
| 4. Percentage in Software Engineering       | This feature contains percentile in Software Engineering. |
| 5. Percentage in Computer Networks          | This feature contains percentile in Software Engineering. |

|                                        |                                                             |
|----------------------------------------|-------------------------------------------------------------|
| 6. Percentage in Electronics Subjects  | This feature contains percentile in Computer Networks.      |
| 7. Percentage in Computer Architecture | This feature contains percentile in Computer Architecture.  |
| 8. Percentage in Mathematics           | This feature contains percentile in Mathematics.            |
| 9. Percentage in Communication skills  | This feature contains percentile in Communication skills.   |
| 10. Hours working per day              | This feature contains no. of working hours an employee has. |
| 11. Logical quotient rating hackathons | This feature contains students rank in hackathons.          |
| 12. Coding skills rating               | This feature contains students rate in coding.              |
| 13. Public speaking points             | This feature contains feedback on students speaking.        |
| 14. Can work long time before system   | This feature contains percentile in Communication skills.   |
| 15. Self-learning capability           | This feature contains student's self learning rate.         |
| 16. Extra-courses did                  | This feature contains percentile in                         |

|                                                 |                                                                                   |
|-------------------------------------------------|-----------------------------------------------------------------------------------|
|                                                 | Communication skills.                                                             |
| 17. Certifications                              | This feature contains all students certifications.                                |
| 18. Workshops                                   | This feature contains all workshops student attend.                               |
| 19. Talent tests taken?                         | This feature contains whether a student taken talent tests.                       |
| 20. Olympiads                                   | This feature contains students olympiads and their scores.                        |
| 21. Reading and writing skills                  | This feature contains students reading and writing ability.                       |
| 22. Memory capability score interested subjects | This feature contains memory capability score of interested subjects for student. |
| 23. Interested career area Job/Higher studies?  | This feature contains student interest career area Job/Higher studies.            |
| 24. Type of company want to settle in ?         | This feature contains student interest type of company they want to settle in.    |
| 25. Taken inputs from seniors or elders         | This feature contains inputs from seniors or elders for student.                  |
| 26. Interested in games                         | This feature contains students games that he or                                   |

|                                         |                                                                     |
|-----------------------------------------|---------------------------------------------------------------------|
|                                         | she likes to play.                                                  |
| 27. Interested type of subjects         | This feature contains subjects student is interested in.            |
| 28. Salary range expected?              | This feature contains salary range student expects.                 |
| 29. Gentle or tuff behaviour            | This feature contains students behaviour.                           |
| 30. Management or technical salary/work | This feature contains students Management or technical salary/work. |
| 31. Hard/Smart worker                   | This feature contains what type of worker is the student.           |
| 32. Worked in teams ever?               | This feature contains the experience of student working in teams.   |
| 33. Introvert                           | This feature contains whether the student is introvert or not.      |
| 34. Suggested job role                  | This feature contains the result i.e., suggested job.               |

mark in range 60%-80% is class 'Medium' and mark below 60% is 'Low'. Data processing is essential to recover the eminence of the dataset. Text data type attributes Gender, Parent Responsible, Parent Answering, Parent Satisfaction, Student Absence Days are transformed to binary data '0' and '1'. Other Text data type attributes Nationality, Place of Birth, Stage, Grades,

SectionID, Topic, Class in 1st Term and Class in 2nd Term are transformed to numerical data type

#### IV. METHODOLOGY

In this paper, a modified vanilla RNN is used to make the prediction of a student's performance. After building the proposed system, a comparative study is made with the Deep neural network (DNN) which is proposed by the Prabu P, et al.[14]. The whole implementation is done in PYTHON. Once the data pre-processing is completed, the whole data set is classified into two parts. One is training set and another one is the testing set. The data set is divided into the ratio of 3:1(Train/Test). The class variable is one-hot encoded for both train and test dataset. The class variable is represented in the numerical format as below

| CLASSES | One-HotEncodingFormat |
|---------|-----------------------|
| Low     | [1,0,0]               |
| Medium  | [0,1,0]               |
| High    | [0,0,1]               |

The proposed neural network contains three layers. The first layer is the input layer and the seventeen features are given as input in the system. The final result of the student is produced as output from the output layer. The hidden layer is located in between the input and the output layer. The hidden layer consists of 40 neurons. One recursive path is there in between the hidden layer neurons to make the neural network recurrent. Random initialization of weights  $w$  and bias  $b$  to every interconnected layers. This initialization is specifically proposed for hidden nodes with Rectified Linear Unit (ReLU) non linearity. The ReLU activation function is defined as the positive part of its argument. It is represented by

$$f(x) = x + \max(0, x) \quad \text{Where } x \text{ is the input to a neuron.}$$

## V. RESULT

The proposed method is applied to a dataset of 480 students. 360 data is used for training and the rest 120 data is used for testing. The proposed RNN is able to achieve the accuracy of 85.4% in predicting of a student's final grade. The performance of the training data is determined using the mean square error. After completing the 200 epochs, the mean square error of the training data set is determined as 0.11.

### ClassificationResults:-

| CLASSIFIER                | ACCURACY (%) |
|---------------------------|--------------|
| Artificial Neural Network | 77.8         |
| Deep Neural Network       | 83.7         |
| Recurrent Neural Network  | 85.4         |

## VI. CONCLUSION

Here a method is presented to forecast a student presentation utilizing Recurrent Neural Network with Rectified Linear Unit (ReLU) initiation purpose. RNN is suitable for analysis of the time series data. In this paper, student's ultimate term class is projected utilizing first two term class along side by other fifteen features. Here it is assumed that the student's final class is a sequential output of other two term class. The projected ideal achieved an exactness of 85.4%, that outperformed the other machine learning algorithm. Higher accuracy can be obtained with larger dataset and features. This model can be used to predict student's performance and help to identify the students who have higher chance of failing. As a part the future work, the

model can be tested with larger data set by adding more features. A clustering technique can be used to categorize the students at the very early phase, which will help to deal with the cold start problem. The proposed method can also be used in the intelligent tutorial system (ITS) to predict a student performance and based on that ITS can provide proper feedback to the students.

## REFERENCES

- [1]. P. Saini , A.K. Jain, "Prediction using Classification Technique for the Students' Enrollment Process in Higher Educational Institutions", International Journal of Computer Application, Springer, Berlin, Heidelberg, 2013, 84–89.
- [2]. M. Koutina, K.L. Kermanidis, "Predicting Postgraduate Students' Performance using Machine Learning Techniques", in Artificial Intelligence Applications and Innovations, Springer, Berlin, Heidelberg, 2011, 159– 168.
- [3]. H. Agrawal , H Mavani, "Student Performance Prediction using Machine Learning", in International Journal of Engineering Research and Technology, 2015, 271–280.
- [4]. Y. LeCun, Y. Bengio and G. Hinton, "Deep Learning", Nature, 521(7553), 436–444. [
- [5]. G.E. Dahl, D. Yu, L. Deng and A. Acero, "ContextDependent Pre-trained Deep Neural Networks for LargeVocabulary Speech Recognition", IEEE Transactions on audio, speech and language processing, 20(1), 2012, 30–42.
- [6]. R. Collobert , J. Weston, "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning", in proceedings of the 25th International Conference on Machine Learning, 2008, 160–167.
- [7]. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and T. Darrell, "Convolution Architecture for Fast



- Feature Embedding”, in proceedings of the 22nd ACM International Conference on Multimedia, 2014, 675–678.
- [8]. E. Kyndt, M. Musso, E. Cascallar and F. Dochy, “Predicting Academic Performance in Higher Education: Role of Cognitive, Learning and Motivation,” Earli Conference, 2011.
- [9]. Livieris, et al, “Predicting Students’ Performance using Artificial Neural Networks”, 8 th PanHellenic Conference with International Participation Information and Communication Technologies, 2012, 312–328.
- [10]. S. Kotsiantis, et al, “Preventing Student Dropout in Distance Learning Systems using Machine Learning Techniques”, Applied Artificial Intelligence, 18(5), 2003, 411–426.
- [11]. C.E. Moucary, M. Khair and W. Zakhem, “Improving Student’s Performance using Data Clustering and Neural Networks in Foreign-Language Based Higher Education”, The Research Bulletin of Jordan ACM, 2(3), 2011, 27–34.
- [12]. P. Mukta, A. Usha, “A Study of Academic Performance of Business School Graduates using Neural Networks and Statistical Techniques”, Expert Systems with Applications, vol. 36, 2009, 7865–7872.
- [13]. E.A. Amrieh, T. Hamtini and I. Aljarah, “Mining Educational Data to Predict Student’s Academic Performance using Ensemble Methods”, International Journal of Database Theory and Application, 9(8), 2016, 119–136.
- [14]. Prabu P. , Bendangnuksung, “Students’ Performance Prediction using Deep Neural Networks”, International Journal of Applied Engineering Research, vol. 13, 2018, 1171–1176.