Using Deep Learning to Predict Scholarship Scheme Based on Student Details

A. Alsameema, E. Anupriya, M. R. Elakhia, B. Jaya Meera, S. Maheswari

Abstract: The government launches various ambitious programs to make the country more prosperous, but they fail in successful implementation. The main reason behind this issue is the lack of awareness among rural people. This project is to provide a solution to this unaware situation. Through this system, rural students will get to know about what are the various schemes that are furnished by the government. Initially, this system will explore government schemes that are available for the welfare of rural students. Next, the student's dataset (i.e.) name, age, caste, occupation, annual income, etc.) are collected. Then both the datasets are imported into the Anaconda Navigator. Later analysis and classification are done based on communities (SC, ST, BC MBC, and DNC), Educational category (Pre-metric/Post-metric), Board of education (Government/Government-aided), Day scholar or hosteller, age of the students and the schemes are predicted.

Keywords: Recurrent Neural Network(RNN) and Long Short term Memory(LSTM).

I. INTRODUCTION

Nowadays, the government launches various ambitious programs trying to make the country more prosperous, but what they fail is in successful implementation and reaching to beneficiaries. The main reason behind this issue is the lack of awareness among rural people. This paper is to provide a solution to this unaware situation. Through this system, rural students will get to know about what are the various schemes that are furnished by the government. If the rural students came to know about the scholarship schemes and also if they become aware of all the government schemes provided by the Government of India for the welfare of the rural students, then their life would step into the next level. Initially, this system will explore the available government schemes in education for the welfare of rural students. Next, the student’s dataset (i.e.) name, age, caste, occupation, annual income, etc.) are collected. Then both the datasets are imported into the Anaconda Navigator. Later analysis and classification are done based on communities (SC, ST, BC MBC, and DNC), Educational category (Pre-metric/Post-metric), Board of education (Government/Government-aided), Day scholar or hosteller, age of the students and the schemes are predicted. By using Long Short Term Memory (LSTM) algorithm in Recurrent Neural Network(RNN), the scholarship scheme is predicted for the respective students.

Thus, through this project, the rural students will come to know about various beneficiary schemes provided by the government.

II. EXISTING SYSTEM

Alistair Johnson, Harini Suresh, Leo Anthony Celi, MarzyehGhassemi, Nathan Hunt, and Peter Szolovits, in [1] proposed “Clinical Intervention Prediction and Understanding using Deep Networks” in which data from all available ICU sources and focus on learning rich representations of this data to predict onset and meaning of multiple invasive interventions.

Arindam Mondal and Joydeep Mukherjee, in [2] proposed "An Approach to Predict a Student's Academic Performance using Recurrent Neural Network (RNN)," in which they predict student's final grades.

Atsushi SHIMADA, Fumiya OKUBO, Shinichi KONOMI, and Takayoshi YAMASHITA, in [3] proposed "Student's Performance Prediction Using Data of Multiple Courses by Recurrent Neural Network" in which they predict student's final grades.

Giuseppe Serra, Lorenzo Baraldi, in [4] Marcella Cornia, and Rita Cucchiara, proposed "Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model" in which a novel saliency prediction architecture that incorporates an Attentive Convolutional Long Short-Term Memory network that iteratively focuses on appropriate spatial locations to refine saliency features.

Hangzhou Yang, and Kuang Junwei, in [5] proposed “Dynamic prediction of cardiovascular disease using improved LSTM” in which Previous dynamic prediction models rarely handle multi-period data with different intervals, and the large-scale patient hospital records.

Kamurunnissa Bee, K., Kiruthika, J., and Sowri Raja Pillai., in [6] "predicted Prediction Of Heart Disease Using RNN Algorithm" in which RNN Classifier is used as ensemble to incorporate different base.

Karan Manchanda and NavdeepKhare, in [7] proposed “Implementation of student performance evaluation through supervised learning using neural network,” which gives solution for the problem by designing a user interface which work on supervised learning using Neural Network.

Lia Bally., Marko Jankovic, V., Qingnan Sun., and StavroulaMougiakakou, G., in [8] proposed “Predicting Blood Glucose with an LSTM and Bi-LSTM Based Deep Neural Network” in which deep learning network was used to predict future blood glucose levels.
Using Deep Learning to Predict Scholarship Scheme Based on Student Details

Lokeswari, Y., V., RaghavNandakumar., Uttamraj, K., R., in [9] and Vishal, R., proposed "Stock Price Prediction Using Long Short Term Memory" in which an online learning algorithm for predicting stock price.

Poornima, S., and Pushpalatha, M., in [10] proposed “Prediction of Rainfall Using Intensified LSTM Based Recurrent Neural Network with Weighted Linear Units” in which they uses RNN algorithm to predict rainfall.

III. PROPOSED SYSTEM

We create a dataset using the details of the scholarship schemes collected. The scholarship scheme dataset consists of attributes like Scheme name, Educational Category, Board of Education, Attempts, Parent's Occupation, Parent's Income, DayScholar/Hosteller, Extra preference for Disability, Extra preference for Merit Students, Extra preference for Girls, Scholarship Amount, Community. Next, we’ll gather the student’s data. The student dataset consists of attributes like Name, Age, Educational Category, Board of Education, Community, Gender, Parent’s Occupation, Parent’s Income, Dayscholar or Hosteller, Physically challenged, Merit student and Attempts. Then both the datasets are imported into the Anaconda Navigator. Then using the recurrent neural network, we will predict the schemes for the particular student. In the recurrent neural network, the datasets are classified as training and testing datasets. 80% of datasets are used for training, and the remaining 20% of datasets are used for testing. Thus, through this project, the rural students will come to know about various beneficiary schemes provided by the government, and they can view the scholarship schemes available for them. They can utilize those schemes for the improvement of rural surroundings.

The student dataset was collected from government and government-aided schools. The student dataset consists of 13 attributes and 827 instances. While, the schemes dataset consists of the attributes such as scheme name, educational category, the board of university, attempts, parent's occupation, parent's income, day scholar hosteller, other preference for disability, extra preference for merit students, extra preference for girls, scholarship amount and age. The schemes dataset consists of 13 attributes and 100 instances.
layers, batch size, etc. The batch size is set as 64, and the epoch is set as five, which iterates the dataset.

3.3 Module 3: Finding Accuracy

Root Mean Square Error (RMSE): Measure of the difference between values predicted by a model and the values observed. It is calculated by taking the summation of the squares of the differences between the predicted value and actual value and dividing it by the number of samples. It is mathematically expressed as follows:

\[
RMSE = \sqrt{\frac{\sum (Y_{\text{test}} - Y_{\text{predict}})^2}{N}}
\]

\(Y_{\text{test}}\) is the actual observation, and \(Y_{\text{predict}}\) is the predicted observation.

\(N\) is the number of samples.

Mean Square Error = 0.153532

Root Mean Square Error = \(\sqrt{0.153532} = 0.39183274637860055\)

The accuracy of the prediction model can then be estimated using the RMSE (Root Mean Squared Error). In our project, the LSTM model gave a RMSE value of 0.39183274637860055 from this value we can say that RMSE value ranges from 0 to 1 and lower values of RMSE (closer to 0) indicate better fit.

IV. IMPLEMENTATION

4.1 Dataset Collection

The first dataset is the student dataset which contains the attributes such as name, age, educational category, board of university, community, gender, parent's occupation, parent's income, day scholar hosteller, physically challenged, merit student and attempts (the number of attempts a particular student take to pass a class). The student dataset consists of 13 attributes and 827 instances. The second dataset is the scholarship scheme dataset which consists of the attributes such as scheme name, educational category, the board of university, attempts, parent's occupation, parent's income, day scholar hosteller, extra preference for disability, extra preference for merit students, extra preference for girls, scholarship amount and age. The scholarship schemes dataset consists of 13 attributes and 100 instances.

4.2 Encoding Categorical Features

We used a label encoder for encoding certain features class sklearn.preprocessing.LabelEncoder - Encode target labels with value between 0 and n_classes-1. In our project, we have encoded attributes like educational category, the board of education, parent's income, extra preference for disability, extra preference for merit student, extra preference for girls, community, day scholar/hosteller in both student and scholarship scheme dataset. Label encoder transforms non-numerical labels (as long as they are hashable and comparable) to numerical labels.

4.3 Dataset Splitting

In this project, we have split the student dataset into train and test dataset using train_test_split library function. The training set should be a random selection of 80% of the original data. The testing set should be the remaining 20%. The student dataset consists of 827 instances, and it splits into the train and test dataset. The training dataset consists of 661 instances, and the test dataset consists of 166 instances.

Testing dataset = 827*(20/100) = 166 instances

Training dataset = 827*(80/100) = 661 instances

4.4 Predicting Scholarship Scheme With Lstm

To build the LSTM, we need to import a couple of modules from Keras:

- Sequential for initializing the neural network
- Dense for adding a densely connected neural network layer
- LSTM for adding the Long Short-Term Memory layer
- Dropout for adding dropout layers that prevent overfitting

In our project we combine the LSTM layer and later add a few Dropout layers to avoid overfitting. We combine the LSTM layer with the following arguments: 100 units, which is the dimensionality of the output space. The Figure 5.5 shows the output using the LSTM model and the scholarship scheme and scholarship amount available for the student K.Kavitha when the user enters the serial number of the student. It will display the Educational category, Board of Education, Dayscholar/Hosteller, community, and age for the given serial number of the student. The LSTM model outputs predicted scheme for the given student K.Kavitha would get Pre Matric Scholarship for BC/MBC/DNC students. Predicted scheme amount for the given student K.Kavitha will get 200 per month.

V. RESULTS

Figure 5.1 Student Dataset

Figure 5.2 Scholarship Scheme Dataset
VI. CONCLUSION

Our Indian government inaugurates numerous ambitious programs to make the country more harmonious and prosperous. Through this innovative proposal, the utilization of government schemes will be increased among rural people. The rural people can effectively make use of all the projects furnished by the government of India for their welfare to the maximum extent. The amount spent by the government for the welfare schemes will yield a worthy outcome. Thus according to the saying, “Developed Villages—Developed Nation,” India will move towards a prosperous future.
REFERENCES

1. Alistair Johnson., Harini Suresh., Leo Anthony Celi., MarzyehGhassemi., Nathan Hunt., and Peter Szolovits., 'Clinical Intervention Prediction and Understanding using Deep Networks' arXiv preprint arXiv:1705.08498 (2017).

2. Arindam Mondal., and Joydeep Mukherjee., 'An Approach to Predict a Student's Academic Performance using Recurrent Neural Network (RNN),' Corpus ID:70189807.0a

3. Atsushi SHIMADA., Fumiya OKUBO., Shinichi KONOMI., and Takayoshi YAMASHITA., 'Students Performance Prediction Using Data of Multiple Courses by Recurrent Neural Network,' Chen, W. et al. (Eds.) (2017).Proceedings of the 25th International Conference on Computers in Education.

4. Giuseppe Serra., Lorenzo Baraldi., Marcella Cornia., and Rita Cucchiara., 'Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model.'

5. Hangzhou Yang., and Kuang Junwei., 'Dynamic prediction of cardiovascular disease using improved LSTM,' International Journal of Crowd Science.

6. Kamurunnissa Bee, K., Kiruthika, J., and Sowri Raja Pillai., 'Prediction Of Heart Disease Using RNN Algorithm,' Prediction Of Heart Disease Using RNN Algorithm.

7. Karan Manchandia., and NavdeepKhare., 'Implementation of student performance evaluation through supervised learning using neural network,' International Journal of Engineering Sciences & Research Technology, [Manchandia et al., 6(3): March 2017]

8. Lia Bally., Marko Jankovic, V., Qingnan Sun., and Stavroula Mougiakakou, G., 'Predicting Blood Glucose with an LSTM and Bi-LSTM Based Deep Neural Network'

9. Lokeswari Y. V., RaghavNandakumar., Uttamraj, K. R., and Vishal, R., 'Stock Price Prediction Using Long Short Term Memory' International Research Journal of Engineering and Technology (IRJET) Volume:05 Issue: 03|Mar-2018.

10. Poornima, S., and Pushpalatha, M., 'Prediction of Rainfall Using Intensified LSTM Based Recurrent Neural Network with Weighted Linear Units.'