Predicting Student Exam Scores Using Hybrid Neural Networks Incorporating BERT and Categorical Features

Names:

* Hamza Mohmmed Elshennawy 1710874
* Omar Abdelrhman Abdelaziz Esmail 221000369
* Mazen Elsayed 202000782

Under the supervision of Dr. Ghada Khoriba.

## Abstract

This study presents a comparative analysis of two neural network models for predicting student exam scores. The first model employs a hybrid architecture, integrating BERT embeddings to extract semantic insights from teacher notes and dense layers to process categorical features such as motivation levels and parental involvement. The second model, a simpler design, focuses solely on structured categorical data. Trained on a dataset of 6,379 student records, the hybrid model achieved a remarkable accuracy of 90%, while the simpler model attained a modest accuracy of 65%. These results underline the effectiveness of multi-modal approaches in educational analytics and highlight the limitations of relying solely on structured data.

## 1. Introduction

Predicting student performance is a critical task in educational research aimed at enabling personalized learning and timely interventions. Traditional models often rely on structured data, such as demographic information, grades, and attendance. However, they fail to harness the wealth of unstructured textual data, such as teacher notes, which can provide deep contextual insights into student behavior and challenges.  
This study introduces two predictive models to address this gap. The first model integrates the advanced capabilities of BERT (Bidirectional Encoder Representations from Transformers) for processing textual data with dense layers for structured categorical features. The second model focuses solely on structured data, serving as a baseline for comparison. By evaluating the performance of these two models, this paper demonstrates the superior efficacy of hybrid architectures in educational analytics.

## 2. Related Work

Numerous studies have explored the integration of textual and structured data in predictive modeling:  
- **Textual Data in Education:** Sun et al. (2020) highlighted the importance of analyzing textual feedback, such as teacher comments, to predict student engagement and performance.  
- **Hybrid Neural Networks:** Xu et al. (2021) proposed a hybrid architecture combining text embeddings with structured features, improving predictions for student dropout rates.  
- **Transformer Models:** The introduction of BERT by Devlin et al. (2018) revolutionized natural language processing, providing state-of-the-art results in tasks such as sentiment analysis and classification. Its adaptability to educational datasets has been demonstrated by subsequent studies.  
Building on this foundation, our work extends the application of hybrid neural networks by integrating BERT with categorical data processing for predicting student exam scores.

## 3. Methodology

### 3.1 Dataset

The dataset comprises 6,379 student records, featuring:  
- **Teacher Notes:** Descriptive textual data on student performance and behavior.  
- \*\*Motivation Level\*\*: Categorical feature indicating the student’s self-reported motivation.  
- **Parental Involvement:** Categorical feature reflecting the degree of parental engagement in education.  
- **Exam Scores:** Continuous variable representing student exam performance, used as the prediction target.

### 3.2 Preprocessing

**Textual Data**  
- Tokenized using BERT’s tokenizer, ensuring compatibility with the pretrained BERT model.  
- Input sequences were truncated or padded to a maximum length of 128 tokens.  
- Converted into input IDs and attention masks for BERT’s processing pipeline.  
  
**Categorical Data**  
- One-hot encoding was applied to categorical features (motivation levels and parental involvement) using a scikit-learn pipeline, ensuring efficient representation for neural network input layers.

### 3.3 Model Architectures

**Hybrid Neural Network:**  
1. **BERT Layer:** Pretrained BERT (bert-base-uncased) extracts semantic embeddings from textual data. The output of the [CLS] token is pooled to represent the context of teacher notes.  
2. **Dense Layer for Categorical Data:** Processes one-hot encoded features, projecting them into a 32-dimensional latent space.  
3. **Fusion Layer:** Concatenates the outputs from the BERT layer and the categorical dense layer, passing them through a fully connected layer with 64 units.  
4. **Output Layer:** A regression head predicts the exam score based on the combined feature representation.  
  
**Simple Neural Network:**  
1. **Dense Layers:** Processes only the one-hot encoded categorical features through multiple dense layers.  
2. **Output Layer:** A regression head predicts the exam score.

## 4. Results

### 4.1 Experimental Setup

- **Data Split:** The dataset was divided into training (70%), validation (15%), and test (15%) subsets.  
- **Loss Function:** Mean Squared Error (MSE) was employed to evaluate prediction accuracy.  
- **Optimizer:** Adam optimizer with a learning rate of 0.0001 was used for both models.  
- **Batch Size:** A batch size of 32 was maintained across experiments.

### 4.2 Performance Metrics

|  |  |  |
| --- | --- | --- |
| Metric | Hybrid Model | Simple Model |
| Training Loss | 78.37 | 215.76 |
| Validation Loss | 14.91 | 57.84 |
| Accuracy | 90% | 65% |

The hybrid model consistently outperformed the simpler model, achieving superior accuracy and lower validation loss. The incorporation of BERT embeddings provided nuanced contextual understanding, significantly enhancing predictive performance.

# 5. Discussion:

The study highlights the advantages of hybrid architectures in leveraging both textual and categorical data for educational analytics. The BERT layer effectively captures semantic nuances from teacher notes, while the dense layers encode structured categorical features. This multi-modal approach provides a comprehensive representation, enabling more accurate predictions.

In contrast, the simpler model, limited to categorical data, achieved moderate accuracy but lacked the depth required for complex predictive tasks. These findings underscore the potential of integrating transformer-based models like BERT into hybrid architectures for educational applications.

Future work could explore:

* Incorporating additional data modalities, such as attendance records and peer interactions.
* Experimenting with alternative transformer models, including RoBERTa and DistilBERT.
* Extending the model to predict other educational outcomes, such as course completion rates or overall academic performance.

# 6. Conclusion:

This study introduced two neural network models for predicting student exam scores. The hybrid model, integrating BERT embeddings and categorical feature processing, achieved an impressive accuracy of 90%, significantly outperforming the simpler model, which achieved 65%. These results validate the effectiveness of multi-modal approaches in educational analytics, providing a robust foundation for future research.

# References:

1. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
2. Sun, Z., Wang, S., Li, M., & Liu, J. (2020). Using textual feedback to predict student engagement: A case study. *Journal of Educational Data Science*, 7(3), 45-57.
3. Xu, Y., Zhang, W., & Zhao, H. (2021). Hybrid neural networks for student performance prediction: Text meets structured data. *Proceedings of the International Conference on Educational Data Mining*.