

# Capstone Project Weekly Report

**Date:** 6/10/2024

## Project Details:

1. Sponsor Company: Prepshala.Pvt.LTd
2. Project Title: An adaptive screening tool using Generative AI and LLMs for the assessment of language and personality traits

**Note:** All the fields in the form are required.

## Project Milestones:

Progress made in current week and contribution from individual team members:

During the current week, significant progress was achieved in developing the assessment model utilizing the Llama architecture. Within the first two days, we successfully built and demonstrated the model to the project sponsor. We gathered valuable feedback from the sponsor mentor, particularly regarding the reduction of essay length and the need to develop a model capable of predicting IELTS bands based on responses ranging from 5 to 10 lines.

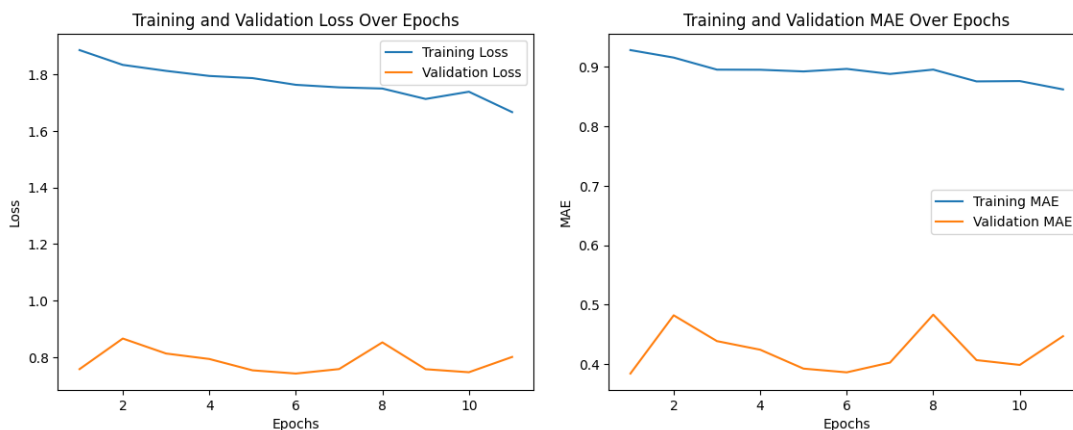
In addition to demonstrating the model, we outlined our approach for the writing module. Specifically, we proposed that the first two questions would be of an easy level and not evaluated, while the remaining three questions would be of medium and difficult levels. We highlighted the reasoning behind this approach.

Sponsor provided valuable feedback, suggesting that we should use prompts or questions from the existing dataset to ensure credibility and reliability in the model's predictions. We adopted this suggestion and incorporated it into the models we developed.

Following the feedback, we consolidated all unique records from the dataset and applied a summarization model to streamline the data. Subsequently, we developed a deep neural network model to predict IELTS bands, achieving promising results with a Mean Absolute Error (MAE) of 0.5. After experimenting with two models, we finalized one that effectively addressed the issue of overfitting.

All team members actively contributed to the model development and refinement process.

Here are our model results:



```

y_pred = model.predict(X_val).flatten() # Flatten to get predictions in a 1D array
y_actual = y_val.flatten() # flatten the actual values as well for consistency

# Create a DataFrame to store Actual vs Predicted values
df_comparison = pd.DataFrame({'Actual Band': y_actual, 'Predicted Band': y_pred})

# Display the first few rows of the comparison DataFrame
print(df_comparison)

```

	Actual Band	Predicted Band
0	6.5	6.537517
1	7.5	7.523667
2	7.0	6.729431
3	7.5	6.945408
4	9.0	8.667498
...	...	...
696	8.5	7.484505
697	6.5	6.788659
698	5.5	5.242335
699	4.0	3.924852
700	6.0	5.520251

[781 rows x 2 columns]

```

pred = model.predict(response)
predicted_scores.append(pred[0][0])
return predicted_scores

# Predict band scores for dataset-based responses
predicted_band_scores = predict_band_scores(model, tokenized_responses)

# Display predicted band scores
for i, score in enumerate(predicted_band_scores):
    print(f"Response {i+1} (Dataset Question) predicted band score: {score}")

```

Prompt (Not Evaluated): The ocean is \_\_\_\_\_. (Please fill in the blank with \*\*one word\*\*.)  
Your response: blue  
Prompt: Nowadays more and more people want to get things done instantly (services, information, tasks). Why is this? Do you think it is a positive or negative de  
Your response: The world is evolving rapidly so does people mind. Everyone need instant results and success. Services information and tasks speed has improved raj  
Prompt: Some people think that in the modern world we are more dependent on each other while other think that people have become more independent.discuss both vi  
Your response: Most of the people are trying to be independent in the modern world. It has both pros and cons. Being more independent at times might not achieve  
WARNING:absl:Compiled the loaded model, but the compiled metrics have yet to be built. 'model.compile\_metrics' will be empty until you train or evaluate the mode.  
1/1 ----- 0s 190ms/step  
1/1 ----- 0s 16ms/step  
Response 1 (Dataset Question) predicted band score: 5.653146266937256  
Response 2 (Dataset Question) predicted band score: 5.440123081207275

## Tasks to finish in next week:

- Model Demonstration and Feedback:**  
Present the current model to the project sponsor for feedback. Based on the inputs received, implement necessary adjustments to enhance the model's performance.
- Connect with Project Mentor:**  
Schedule a meeting with the project mentor to demonstrate the progress made so far and obtain their valuable inputs for further refinement.
- Research on Deployment and UI Framework:**  
Conduct research into potential deployment platforms and UI frameworks that will support the model's integration.
- Initiation of Listening Module Research:**  
If the model is finalized, start research on the next project phase—the Listening Module. Investigate methodologies and frameworks necessary for evaluating listening skills effectively.

## **Updates/MoM from Sponsor and Faculty Mentor:**

**Note:** It is expected that you have at least one weekly connect with the faculty mentor and sponsor. If you were not able to schedule meetings with the sponsor or faculty mentor in the current week, please mention the reason for your inability to meet with the Sponsor or Faculty Mentor.

### **Updates/MoM from Sponsor:**

On 2nd October 2024, the team held a 45-minute Zoom meeting with the project sponsor to present the progress made on the model development. The sponsor provided positive feedback, commending the team's efforts. During the session, several key inputs were shared to further refine the model.

The sponsor requested modifications to the dataset, specifically suggesting the use of shorter essays (5-10 lines) for evaluation, as the current dataset contained longer essays. In response, the team was tasked with revising the dataset to reflect these changes.

Additionally, we outlined our approach for the writing module during the meeting. We proposed that the first two questions be of an easy level and not evaluated, while the remaining three questions would be of medium and difficult levels. The reasoning behind this approach was explained, highlighting that it would help ease students into the assessment before transitioning to more challenging questions.

The sponsor provided constructive feedback on this, recommending that we use prompts and questions from the existing dataset to ensure credibility and reliability. We have adopted this suggestion in the models we developed.

Following the sponsor's feedback, the team conducted research on various text compression techniques and concluded that applying a summarization model was the most effective approach. Consequently, we applied the summarization model to the unique dataset, which consisted of 3,506 records. After implementing this step, we ran the deep neural network model, achieving a Mean Absolute Error (MAE) of 0.5.

### **Updates/MoM from Faculty Mentor:**

Our project charter was reviewed by the faculty mentor, faculty responded positively over an email, expressing approval of the overall direction outlined in the document.

Due to time constraints and ongoing commitments with office work and model development, we were unable to schedule a meeting with the faculty mentor this week. However, in the upcoming week, we plan to demonstrate the current state of the model to the faculty mentor, obtain their inputs, and make the necessary modifications based on their feedback.

## **Challenges:**

Mention any technical and non-technical challenges that you faced during the current week that hindered your project progress. Enter "NA" if you didn't face any challenges.

### Technical Challenges:

1. **Dataset Modification:**

One of the primary challenges we encountered was that all available datasets contained full-length IELTS standard essays, which were much longer than the required 5-10 lines. To address this, we applied a summarization model to shorten the existing essays in the dataset, enabling us to proceed with the model training as per the sponsor's requirements.

2. **Model Overfitting:**

One of the two models we developed exhibited significant overfitting during testing. We resolved this issue by implementing various hyperparameter tuning techniques, applying regularization methods, and utilizing cross-fold validation. As a result, we were able to improve the model's performance, achieving a Mean Absolute Error (MAE) of 0.5.

3. **Access to Llama Model:**

We successfully utilized the open-source Llama model via a free API, which enabled us to integrate a large language model into our project for free.

### Non-Technical Challenges:

NA

Mention any other queries/challenges regarding the project that you want to highlight:

NA

